Stable Takens’ Embeddings for
Linear Dynamical Systems

Han Lun Yap and Christopher J. Rozell

October 2010

Abstract

Takens’ Embedding Theorem remarkably established that concatenating \( M \) previous outputs of a dynamical system into a vector (called a \textit{delay coordinate map}) can be a one-to-one mapping of a low-dimensional attractor from the system state-space. However, Takens’ theorem is fragile because even small imperfections can induce arbitrarily large errors in the attractor representation. We extend Takens’ result to establish explicit, non-asymptotic sufficient conditions for a delay coordinate map to form a \textit{stable embedding} in the restricted case of linear dynamical systems and observation functions. Our work is inspired by the field of Compressive Sensing (CS), where results guarantee that low-dimensional signal families can be robustly reconstructed if they are stably embedded by a measurement operator. However, in contrast to typical CS results, i) our sufficient conditions are independent of the size of the ambient state space \((N)\), and ii) some system and measurement pairs have fundamental limits on the conditioning of the embedding (i.e., how close it is to an isometry), meaning that further measurements beyond some point add no further significant value. We use several simple simulations to explore the conditions of the main results, including the tightness of the bounds and the convergence speed of the stable embedding.

Index Terms

Takens’ Embedding Theorem, Linear Systems, Stable Embedding, Restricted Isometry Property, Compressed Sensing
I. INTRODUCTION

Of the many types of data confronting signal processing researchers, time series data is perhaps one of the most common. While there are many possible ways to analyze a time series, one of the most important tasks in many areas of science and engineering is to characterize (or predict) the state of a dynamical system from a stream of its output data [2], [3]. This type of state identification can be particularly challenging because the internal (possibly high-dimensional) system state $x(t) \in \mathbb{R}^N$ is often only indirectly observed via a one-dimensional time series of measurements produced through an observation function $s(t) = h(x(t))$, where $h : \mathbb{R}^N \rightarrow \mathbb{R}$.

Surprisingly, when the dynamical system has low-dimensional structure because the state is confined to an attractor $\mathcal{M}$ of dimension $d$ ($d < N$) in the state space, Takens’ Embedding Theorem [4], [5] shows that complete information about the hidden state of this system can be preserved in the time series output data $s(t)$. Indeed, many systems of interest do have this type of structure [6], and a variety of algorithms for tasks such as time series prediction and attractor dimension estimation exploit Takens’ result [3]. Specifically, Takens defined the delay coordinate map $F : \mathbb{R}^N \rightarrow \mathbb{R}^M$ as a mapping of the state vector $x(t)$ to a point in the reconstruction space $(\mathbb{R}^M)$ by taking $M$ uniformly spaced samples of the past time series (with sampling interval $T_s$) and concatenating them into a single vector,

$$ F(x(t)) = [s(t) \ s(t - T_s) \ s(t - 2T_s) \cdots \ s(t - (M - 1)T_s)]^T. \quad (1) $$

Takens’ main result [4] (later refined in [5]) states that (under a few conditions on $T_s$ discussed later) for almost every smooth observation function $h(\cdot)$, the delay coordinate map is an embedding$^1$ of the state space attractor $\mathcal{M}$ when $M > 2d$. In other words, despite the state being hidden from direct observation, the topology of the attractor that characterizes the dynamical system can be preserved in the time series data when it is arranged into a delay coordinate map.

In the absence of imperfections such as measurement or system noise, Takens’ result indicates that a delay coordinate map should be as useful for characterizing a system and predicting future time series values as direct observation of the hidden system state. However, in the presence of noise, a one-to-one mapping is not sufficient to guarantee the robustness of any processing performed in the reconstruction

$^1$An embedding is a one-to-one immersion.
The main underlying problem is that while Takens’ theorem guarantees the preservation of the attractor’s topology, it does not guarantee that the geometry of the attractor is also preserved. For example, Takens’ result guarantees that two points on the attractor $\mathcal{M}$ do not map to the same point in the reconstruction space, but there are no guarantees that close points on the attractor remain close under this mapping (or far away points remain far away). Consequently, relatively small imperfections could have arbitrarily large effects when the delay coordinate map is used in applications.

In the signal processing community, recent work has highlighted the importance of the conditioning of measurement operators to ensure the geometry of a low-dimensional signal family is preserved. Consider a signal class $\tilde{\mathcal{M}}$ with intrinsic dimension $d$ residing in $\mathbb{R}^N$ and measurement operator $\tilde{F} : \mathbb{R}^N \rightarrow \mathbb{R}^M$. To preserve the geometry of the signal class $\tilde{\mathcal{M}}$, we require $\tilde{F}$ to be a stable embedding of $\tilde{\mathcal{M}}$, i.e. for all distinct pairs of points $x, y \in \tilde{\mathcal{M}}$,

$$C(1 - \delta) \leq \frac{\|\tilde{F}(x) - \tilde{F}(y)\|_2^2}{\|x - y\|_2^2} \leq C(1 + \delta).$$

(2)

The convergent value $C$ is a scaling constant that could be absorbed into $\tilde{F}$ and the conditioning number $0 \leq \delta < 1$ bounds how much pairwise distances between signals in $\tilde{\mathcal{M}}$ can change when mapped by $\tilde{F}$ (i.e., how near $\tilde{F}$ is to an isometry). The Johnson-Lindenstrauss (JL) lemma [7], [8] gives an example of a stable embedding of a signal class $\tilde{\mathcal{M}}$ consisting of a point cloud of $d = |\tilde{\mathcal{M}}|$ distinct points in $\mathbb{R}^N$. In this result, a random measurement matrix $\tilde{F}$ with $M = O(\log(d))$ rows ensures that (2) holds with high probability for all pairs of points in the point cloud $\tilde{\mathcal{M}}$. Another example is the recent work in the field of compressed sensing (CS) [9], [10], [11], [12], where the canonical results show that similar random matrices $\tilde{F}$ satisfy the Restricted Isometry Property (RIP) with high probability when $M = O(d \log(N/d))$ [13], [14]. The RIP guarantees that (2) holds for all pairs of $d$-sparse signals (i.e., the signal family $\tilde{\mathcal{M}}$ is comprised of signals on the union of all $d$-dimensional subspaces within $\mathbb{R}^N$).

Beyond extending the concept of the JL lemma from a finite point cloud to an infinite signal family, the CS results show the value of stable measurement operators by also making guarantees about efficient and robust signal recovery from these measurements. The notion of a stable embedding has also been extended to other signal models [15], including manifold signal families [16], [17]. The latter can be seen as an extension of Whitney’s Embedding Theorem [18]; while Whitney’s Embedding Theorem ensures
a one-to-one mapping of a manifold $\tilde{M}$ with dimension $d$ for almost any smooth projection function $\tilde{F}$ given that $M > 2d$, the results in [16] further guarantee that (2) holds over this signal family for a given $\delta$ with high probability when $M = O(d \log(N))$ and when $\tilde{F}$ is a random orthoprojector.\footnote{The required number of measurements $M$ in [16] also depends on some properties of the manifold (e.g., the maximum curvature). Clarkson [17] later improved upon $M$ to remove the dependence on the ambient dimension $N$ and reduce the dependence on certain properties of the manifold.}

While the notion of embedding the state of a dynamical system may seem far removed from the CS results, there is actually a close connection. Takens’ Embedding Theorem can be viewed as a special case of Whitney’s Embedding Theorem where the measurement operator $\tilde{F}$ is restricted to forming a delay coordinate map (i.e., $\tilde{F} = F$) and the manifold is taken to be the state space attractor (i.e., $\tilde{M} = M$).

The main contribution of this paper is to further these connections by establishing sufficient conditions whereby the delay coordinate map is a stable embedding of the state space attractor for linear systems with linear observations functions. Specifically, we say that the delay coordinate map $F$ satisfies the Attractor Restricted Isometry Property (A-RIP) if (2) can hold for a sufficiently small conditioning number $\delta$ for all states $x, y$ on the system attractor $M$ (this property will be precisely defined in Section II-B). The main technical result of this paper establishes deterministic, explicit and non-asymptotic sufficient conditions for the A-RIP to hold with a given conditioning. We also explore the meaning of these conditions for characterizing systems via delay coordinate maps. In particular, the results of this exploration are interesting because they contrast with the standard CS results in two important ways: i) the necessary number of measurements scales with the dimension of the attractor $d$ independent of the dimension of the ambient space $N$, and ii) the conditioning of the operator cannot always be improved by taking more measurements, as some system/observation pairs will have a fundamental limit in how well the system geometry can be preserved. While a similar general result for nonlinear dynamical systems is of great interest, this study of linear systems elucidates some of the unique issues that arise when trying to stabilize the embeddings of dynamical systems, helping to pave the way for extensions to nonlinear systems.

II. Background and Related Work

In this section we will briefly review some preliminaries, including a precise statement of Takens’ theorem, attractors of linear systems, and related work in stable embeddings of attractors and manifolds.
A. Linear Systems and Delay Coordinate Maps

Let a dynamical system be defined by the differential equation:

\[ \dot{x}(t) = \Psi(x(t)), \]

where \( x(t) \in \mathbb{R}^N \) is the system state at time \( t \), and \( \Psi : \mathbb{R}^N \rightarrow \mathbb{R}^N \) is a smooth function. As stated earlier, in this paper we will restrict our examination to embeddings of linear dynamical systems where \( \Psi \in \mathbb{R}^{N \times N} \) is a matrix. Before going on, our discussion of these systems will require us to establish a basic notation for complex vector spaces. For \( u = [u_1 \cdots u_N]^T \in \mathbb{C}^N \), we denote the complex variable by \( j \), the (element-wise) complex conjugate by \( u^* \) and the Hermitian transpose by \( u^H = (u^*)^T \).

Given the system matrix \( \Psi \) and the definition of a dynamical system (3), knowing the state at some fixed time \( t_0 \) is equivalent to knowing the path that the system takes to and from that state (called the flow). Classic results in linear systems theory [19] show that the explicit solution for this path is given by a matrix multiplication:

\[ x(t_0 + t) = e^{\Psi t} x(t_0) = \Phi t x(t_0), \]

where \( \Phi = e^{\Psi t} \) is the flow matrix. Note that this solution is valid for positive or negative values of \( t \), describing the flow both forward and backward from time \( t_0 \).

Delay coordinate maps that embed points on the attractor of a dynamical system are intimately connected with the flow of the system approaching that point. In particular, forming a delay coordinate map of a specific point in the state space requires collecting samples of the system flow backward in time from that point at regular intervals \( T_s \). To enable mathematical descriptions of this sampling operation along the flow, we suppress the implicit dependence on the sampling time \( T_s \) and define the compact notation for the flow matrix as \( \Phi = \Phi_{-T_s} \) so that \( x(t - T_s) = \Phi x(t) \). The delay coordinate map \( F \) with \( M \) delays for the case of linear dynamical systems and linear observation functions \( h \in \mathbb{R}^N \) can then be written as a \( M \times N \) matrix such that:

\[
F(x) = Fx = \begin{pmatrix}
    h^T \\
    h^T \Phi \\
    \vdots \\
    h^T \Phi^{M-1}
\end{pmatrix} x.
\]

To ensure that the linear dynamical systems under consideration have non-trivial steady-state behavior
We restrict our study to the class of systems \( A(d) \) described in the following definition.

**Definition II.1.** We say that a linear dynamical system in \( \mathbb{R}^N \) defined by (3) is of Class \( A(d) \) for \( d \leq \frac{N}{2} \) if the system matrix \( \Psi \) is real, full rank and has distinct eigenvalues. Moreover, \( \Psi \) has only \( d \) strictly imaginary conjugate pairs of eigenvalues and the rest of its eigenvalues have real components strictly less than 0. The strictly imaginary conjugate pairs of eigenvalues are called the \( A \)-eigenvalues and they can be expressed as \( \{ \pm j\theta_i \}_{i=1}^d \) where \( \theta_1, \cdots, \theta_d > 0 \) are \( d \) distinct numbers. The corresponding unit norm \( A \)-eigenvectors are \( v_1, v_1^*, \cdots, v_d, v_d^* \). The corresponding eigenvalues of the flow matrix \( \Phi \) are called the \( A_\Phi \)-eigenvalues, and are given by \( \{ e^{\pm j\theta_i T_s} \}_{i=1}^d \).

Moreover, we define \( \Lambda = \text{diag} (j\theta_1, -j\theta_1, \cdots, j\theta_d, -j\theta_d) \) as the diagonal matrix composed of the \( A \)-eigenvalues and \( V = (v_1 | v_1^* | \cdots | v_d | v_d^*) \in \mathbb{C}^{N \times 2d} \) as the concatenation of the \( A \)-eigenvectors into a matrix with \( \text{rank}(V) = 2d \). Since \( \Phi \) is the matrix exponential of \( \Psi \), it is well-known that they share the same eigenvectors [20]. Therefore, if we denote \( D = D_{-T_s} = e^{-\Lambda T_s} \) as the diagonal matrix comprising of the \( A_\Phi \)-eigenvalues, then we have \( \Phi V = V D \).

In order to have a meaningful notion of an embedding, the dynamical system must have its state trajectory confined to a low-dimensional attractor in the state space. Even if the system has transient characteristics from a given starting point, the embedding of this system will only be considered in steady-state when these transients have disappeared. Considering the steady-state dynamics of the system, we make explicit the notion of an attractor through the following definition.

**Definition II.2.** Let a linear dynamical system be of class \( A(d) \) and let \( x_0 = V \alpha_0 \in \mathbb{R}^N \) for some \( \alpha_0 \in \mathbb{C}^{2d} \) be an arbitrary initial state of the system.\(^4\) We define the attractor of this linear dynamical system to be \( \mathcal{M} = \{ x \in \mathbb{R}^N \mid x = V e^{\Lambda t} \alpha_0 , \ t \in \mathbb{R} \} \).

It is easy to see that \( \mathcal{M} \) lives in the span of \( V \). Also, the attractor of the system clearly depends on the initial state of the system. Because the main results of this paper do not depend on the choice of initial state, we will simply refer to the fixed attractor as \( \mathcal{M} \) and suppress the implicit dependence on \( \alpha_0 \).

---

\(^3\)A number \( x \) is strictly imaginary if \( \text{Re}\{x\} = 0 \). This condition ensures that the system modes corresponding to these eigenvectors have persistent oscillation in the steady-state response.

\(^4\)We only need to consider \( x_0 \) in the span of the columns of \( V \) because any orthogonal components vanish in steady-state.
the initial state. Additionally, one can check that this definition meets the basic premise of the notion of an attractor, i.e., that any point on the attractor \( \mathcal{M} \) when projected backwards (or forward) in time by \( \Phi \) will remain on \( \mathcal{M} \). Specifically, for any \( x \in \mathcal{M} \), we can write \( x = V \alpha_x \), where \( \alpha_x = e^{\Lambda t_x} \alpha_0 \) for some \( t_x \in \mathbb{R} \). Then we see that for any \( k \in \mathbb{Z} \), \( \Phi^k x = \Phi^k V \alpha_x = V D^k \alpha_x \), where \( D \) is the diagonal matrix comprised of the \( A_\Phi \)-eigenvalues as defined earlier. First, this means that \( x \) remains on the attractor even when its projected either forward or backward in time by \( \Phi \). Second, we see that in the span of \( v_i \) and \( v_i^* \) for each \( i \), the point \( x \) is proceeding with speed proportional to \( \theta_i T_s \).

**B. Attractor Embeddings**

The following theorem is an extension of Takens’ original result [4], and gives a lower bound on the number of measurements \( M \) sufficient to ensure that a delay coordinate map \( F \) defined as in (1) is a one-to-one mapping from the state-space attractor to the measurement (reconstruction) space.

**Theorem II.1** (Takens’ Embedding Theorem [5]). Assume the dynamical system converges to an attractor \( \mathcal{M} \) of dimension \( d \) and pick a sampling interval \( T_s > 0 \). Let \( M > 2d \) and suppose \( \mathcal{M} \) has a finite number of equilibria, no periodic orbits of \( \Psi \) of period \( T_s \) or \( 2T_s \), and at most finitely many periodic orbits of period \( kT_s \) for \( k = 3, \cdots, M \). Then for almost every smooth function \( h \), the delay-coordinate map \( F \) is one-to-one on \( \mathcal{M} \).

The notion of “almost every” used in the theorem above is technical (see [5] for details), but is consistent with the heuristic notion that out of all possible functions \( h \), most will indeed work.

In this paper we consider the question of when the one-to-one property described in Theorem II.1 can be improved to become a stable embedding where \( F \) is (nearly) an isometry that preserves the geometry of \( \mathcal{M} \). Specifically, we introduce the following definition to formalize the notion of a stable embedding.

**Definition II.3.** We say that a dynamical system in \( \mathbb{R}^N \) that converges to an attractor \( \mathcal{M} \) and a map \( F : \mathbb{R}^N \to \mathbb{R}^M \) satisfy the Attractor Restricted Isometry Property with conditioning \( 0 \leq \tilde{\delta} \leq 1 \) (A-RIP(\( \tilde{\delta} \))) if there exists \( C > 0 \) and \( \delta < \tilde{\delta} \) such that

\[
C(1 - \delta) \leq \frac{\|F(x) - F(y)\|^2}{\|x - y\|^2} \leq C(1 + \delta) \tag{5}
\]

\( ^5 \)In fact, one can show that \( x(t) \) is moving in an elliptical orbit on the span of \( v_i \) and \( v_i^* \) for each \( i \).
holds for all $x, y \in M$.

Note that smaller values of $\tilde{\delta}$ in the above definition imply a more stable embedding because it guarantees that the map is closer to an isometry. Because Taken’s result only guarantees a one-to-one mapping, it is the weakest form of this property, only guaranteeing A-RI P(1).

To see why Takens’ Embedding can be insufficient, we present an illustrative example where the conditioning of the embedding can be made arbitrarily bad when $M$ is the minimum number of delays necessary to satisfy the sufficient conditions of Theorem II.1. Consider a linear system of class $\mathcal{A}(1)$ with $N = 2$, $T_s = 1$, $\mathcal{A}$-eigenvalue $\theta = 0.03$ and $\mathcal{A}$-eigenvector $v = \frac{1}{\sqrt{2}}[1, j]^T$. This system has a circular attractor as depicted in Figure 1(a). The observation function is chosen as $h = \sqrt{\frac{2}{M}}[\sqrt{\epsilon}, \sqrt{1 - \epsilon}]^T$.\(^6\)

Given a particular pair of points $x, y$ on opposite ends of the circular attractor (shown in Figure 1(a)), we examine the ratio $Q(x, y) = \frac{\|F(x) - F(y)\|^2}{\|x - y\|^2}$, where $F$ is the delay coordinate map given in (4). Note that if $F$ is a perfect isometry then $Q(x, y) = 1$, and we must have $Q(x, y) > 0$ for $F$ to be one-to-one. Fixing the number of measurements at $M = 3$ (the minimum required by Takens’ theorem), Figure 1(b) shows the behavior of $Q(x, y)$ for this pair of points as a function of $\epsilon$. We see that while meeting the sufficient conditions of Takens’ Theorem, $\lim_{\epsilon \to 0} Q(x, y) = 0$. Stated another way, by adjusting the parameter $\epsilon$ the conditioning of $F$ can be made arbitrarily bad for this pair of points. To see that this is not simply a bad pairing of the measurement function to the system, note that for any admissible choice of $h$ there would exist a pair of points that would behave the same way.\(^7\) To explore this example further, Figure 1(c) plots $Q(x, y)$ with $\epsilon = 0.1$ and varying $M$ from 3 to 400. We see that with increasing $M$, the ratio $Q(x, y)$ increases, oscillates and converges to a value of $C = 1$. This provides evidence suggesting that as $M$ increases, the conditioning of $F$ improves because the distance between this pair of points is preserved with increasing fidelity. This effect is not predicted by Theorem II.1, but will be shown in our main results in Section III-B.

\(^6\)As will be described in Theorem III.2, the observation function is normalized so that we have an A-RIP convergent value of $C = 1$ regardless of $M$.

\(^7\)One can imagine this by rotating the points $x, y$ by an angle equivalent to the angle between the new measurement function and the given $h$.\(^7\)
Fig. 1. Examination of the conditioning of Takens’ embeddings. (a) The large (blue) circle shows the attractor of the linear system. The (black) diamond and (red) circle markers show 2 different points \(x, y\) that we pick on the opposite ends of the attractor. The arrow depicts the measurement function \(h(\epsilon)\). (b) The graph shows \(Q(x, y)\) for the points \(x, y\) in Figure 1(a) over a range of values of \(\epsilon\) from 0.01 to 0.1. The number of measurements \(M\) is fixed at 3, the minimum required by Takens’ theorem. (c) Here \(Q(x, y)\) is plotted for \(M\) ranging from 3 to 400 (with \(\epsilon\) fixed at 0.1), suggesting a near isometry for \(F\) as \(M\) increases.

C. Related Work

In addition to Takens’ original investigation of attractor embeddings [4], significant advances were made by Sauer et al. [5] to extend these results to include attractors of non-integer dimensions (i.e., strange attractors) and to make the definition of “almost every” more in line with notions of an event that occurs with probability one. Independently but at nearly the same time as Takens’ original work, Aeyels [21] looked at the same problem from a control theory standpoint. He showed that the delay-coordinate map is related to the observability criteria and that given any system in \(N\) dimensions (not just one confined to an attractor), a generic choice of observation function \(h\) guarantees that the system is observable as long as \(M \geq 2N + 1\).

There has also been significant prior work related to embedding manifolds (or fractal sets), which has important implications for attractor embeddings. Specifically, embedding results for manifolds were
derived by Whitney [18] and later expanded on by Sauer et al. [5]. These results show that if a manifold has dimension \(d\), then almost every smooth function mapping into \(\mathbb{R}^M\) with \(M > 2d\) will be an embedding of the manifold. Baranuik et al. [16] extended these results to show that for manifolds with dimension \(d\) embedded in \(\mathbb{R}^N\), random orthoprojections into \(\mathbb{R}^M\) provide a stable embedding of the manifold as long as \(M\) scales linearly with \(d\) and logarithmically with \(N\) (depending also on various properties of the manifold, such as the maximum curvature). Clarkson [17] later improved on the required number of measurements \(M\) by removing the dependence on \(N\) and certain worst case properties of the manifold. We note that these stable embedding results have been used to show that manifold learning and dimensionality estimation algorithms can be performed in the compressed space with nearly the accuracy as they could be performed in the original space [22]. The main distinction between these manifold embedding results and Takens’ theorem is that these results acquire \(M\) independent observations of each single point on the manifold, whereas Takens’ result requires the repeated application of a single observation function to a system having its own internal time variations. In essence, the delay coordinate map relies on the system dynamics to provide measurement diversity when the observations are restricted to a single fixed function \(h\).

One of the principle benefits of a stable delay coordinate map would be resilience to noise and other imperfections. The effect of noise on the reconstruction of state-space attractors has also been previously considered by several researchers apart from the notion of a stable embedding. In [23] the authors looked at a modified embedding theorem for systems corrupted by dynamical noise, considering specifically embeddings using multivariate time series system outputs and taking more measurements than is typically required for a delay-coordinate map. In [24], the authors study the effects of observational noise via statistical methods, showing how the choice of delay-coordinates (i.e., the choice of observation function \(h\) and sampling time \(T_s\) with respect to the system dynamics) affects the ability to make predictions. In particular, they showed that poor reconstruction amplifies noise and increases estimation error.

### III. Stable Embeddings for Linear Dynamical Systems

In this section we present our main technical results. We first present a preliminary result in Section III-A that gives explicit sufficient conditions on the system and observation functions to guarantee that the delay coordinate map is a one-to-one map of the state space attractor. This is akin to Takens’
Embedding Theorem, and we present it here to highlight the specific differences that arise under our restrictions (linear systems and measurement functions) and when seeking explicit conditions on system and measurement pairs (as opposed to the conditions for generic observation functions in Takens’ theorem). We then present our main technical contribution in Section III-B, giving explicit conditions on the system and observation function for the delay coordinate map to be a stable mapping of the attractor where we can make guarantees on the embedding quality and satisfy A-RIP with a specific conditioning number.

A. Takens’ Embeddings

The following theorem gives conditions on the system and the observation function such that the delay coordinate map $F$ is a one-to-one mapping, equivalent to satisfying A-RIP(1). This is analogous to Theorem II.1 in the context of linear dynamical systems and linear observation functions.

**Theorem III.1 (Linear Takens’ Embedding [1]).** Assume a linear dynamical system of class $\mathcal{A}(d)$ in $\mathbb{R}^N$ that is in steady state. Choose $T_s > 0$ to be the sampling interval, $h \in \mathbb{R}^N$ to be the observation function, and denote by $F$ the delay-coordinate map with $M$ delays as defined in (4). Suppose that $M \geq 2d$, the $\mathcal{A}_\Phi$-eigenvalues $\{e^{\pm j\theta_i T_s}\}$ are distinct and strictly complex, and $v_i^H h \neq 0$ for all $i = 1, \ldots, d$. Then for all distinct pairs of points $x, y \in \mathcal{M}$, $F$ satisfies (5) for some constants $C$ and $\delta < 1$.

**Proof:** The proof of this theorem can be found in Appendix A.

To explore the differences that arise in our specific setting of linear systems and linear observation functions, we compare the conditions of this theorem with that of Takens’ theorem. First, we notice that the conditions on the measurement operation are very similar. Theorem III.1 requires $M \geq 2d$, which is a slight improvement over Takens’ $M > 2d$ and likely because of the specific structure of our attractors. There is also a close correspondence with the other condition on the measurement function $v_i^H h \neq 0$. This requirement is an explicit condition on the relationship between the system and observation function ensuring that the observation function can capture some information from every dimension of the attractor.

We note that (Lebesgue) almost-every $h \in \mathbb{R}^N$ will satisfy this condition, and so we find that this is just a more explicit version of Takens’ result that “almost-every” $h$ ensures an embedding.

---

8We say that a number $x$ is strictly complex if $\text{Im}\{x\} \neq 0$. 
Next, we compare our conditions on the system with those imposed by Takens’ theorem. Theorem III.1 requires that the \( \mathcal{A}_\Phi \)-eigenvalues are distinct and strictly complex, which is equivalent to having \( e^{j\theta_p T_s} \neq e^{\pm j\theta_q T_s} \) (distinct) and \( e^{j\theta_p T_s} \neq \pm 1 \) (strictly complex) for all \( p \neq q \) and \( p, q = 1, \ldots, d \). While this requirement implies \(^9\) that \( \mathcal{M} \) does not have periodic orbits of period \( kT_s \) for \( k = 1, \ldots, 2d \) (thus satisfying Takens’ condition), our condition is actually more stringent than this restriction on periodic orbits (likely due to our restricted class of linear observation functions). We note that since \( \{\theta_i\}^d_{i=1} \) are distinct by definition, this condition is dependent on the choice of sampling interval \( T_s \). One can verify that choosing \( T_s < \frac{\pi}{\max\{\theta_i\}} \) is sufficient (but not necessary) to meet the condition of the theorem.

**B. Stable Takens’ Embeddings**

Before presenting our main result giving conditions for a stable embedding of a dynamical system in a delay coordinate map, it will be useful to define and understand the following quantities that characterize how well-behaved the system and measurement process are both individually and jointly. First, we define

\[
\kappa_1 = \min_{i \in \{1, \ldots, d\}} \left\{ \frac{|v_i^H h|}{\|h\|_2} \right\}, \quad \kappa_2 = \max_{i \in \{1, \ldots, d\}} \left\{ \frac{|v_i^H h|}{\|h\|_2} \right\},
\]

characterizing the minimum and maximum projection of the (normalized) observation function on the \( \mathcal{A}_\Phi \)-eigenvectors. Roughly speaking, these quantities are an indication of the disparity between the dimensions of the system attractor that are best and worst matched to the observation function. One would expect that a measurement system is most efficient when it observes all parts of the attractor equally such that \( \kappa_1 \approx \kappa_2 \). Second, we define \( A_1, A_2 \) as the smallest and largest eigenvalues of \( V^H V \), respectively. In a similar vein, these quantities describe how well the system attractor fills the dimensions of the state-space that it occupies (i.e., when \( A_2 \gg A_1 \) the attractor is very elongated in the state space). Again, we would expect that a system will be most amenable to observation when it fills the space such that \( A_1 \approx A_2 \).

Finally, we define

\[
\nu := \max_{p \neq q} \left\{ |\sin(\theta_p T_s)|^{-1}, \left| \sin \left( \frac{(\theta_p - \theta_q) T_s}{2} \right) \right|^{-1}, \left| \sin \left( \frac{(\theta_p + \theta_q) T_s}{2} \right) \right|^{-1} \right\},
\]

\(^9\)This implication can be shown by contradiction. Pick any \( 1 \leq k \leq 2d \) and suppose that \( \mathcal{M} \) has at least a periodic orbit of \( \Psi \) with period \( kT_s \). This would be equivalent to saying that \( e^{j\theta_p k T_s} = (e^{j\theta_p T_s})^k = 1 \) for all \( p \), meaning that for each \( p \) from 1 to \( d \) the quantity \( e^{\pm j\theta_p T_s} \) is uniquely one of the \( k \) roots of unity. However this is impossible as there are \( 2d \) distinct and strictly complex values of \( \{e^{\pm j\theta_i T_s}\} \) and there are only \( k \leq 2d \) roots of unity (including \( \pm 1 \) which are not allowed), and hence we have a contradiction.
which will also bound the constants associated with the stable embedding. Notice that the first term is large if \( \theta_p T_s \) is small for some \( p \) (or that \( \theta_p T_s \approx k\pi \) for some integer \( k \)), meaning that the system state proceeds in the span of \( v_p \) and \( v_p^* \) at a slow pace, thus not producing much diversity in consecutive measurements of the system along these dimensions. The second term is large if \( \theta_p T_s - \theta_q T_s \) is small (or near \( k\pi \)) for some \( p \neq q \) and \( p, q = 1, \ldots, d \), implying that the system state is proceeding in the subspaces spanned by \( v_p, v_p^* \) and \( v_q, v_q^* \) at almost the same rate. This condition would be unfavorable because the system will take an extremely long time to display enough diversity to determine that it is actually traveling on two separate subspaces instead of one. The third term is similar to the second term if we write \( \theta_p T_s + \theta_q T_s = \theta_p T_s - (\theta_q T_s) \). Thus if \( \theta_p T_s \sim -\theta_q T_s \), then the system is again proceeding on two subspaces at almost the same rate (although the system is proceeding in one of the subspaces in the “opposite” direction).

Armed with these definitions, we can now present our main result giving concrete, non-asymptotic guarantees on the conditioning of the delay coordinate map.

**Theorem III.2** (Stable Linear Takens’ Embedding). Assume a linear dynamical system of class \( A(d) \) in \( \mathbb{R}^N \) that is in steady state. Choose \( T_s > 0 \) to be the sampling interval, \( h \in \mathbb{R}^N \) to be the observation function such that \( \|h\|_2^2 = \frac{2d}{M} \), and denote by \( F \) the delay-coordinate map with \( M \) delays as defined in (4). Suppose that \( M > (2d - 1) \frac{A_2\kappa_2^2}{A_1\kappa_1^2} \nu \), the \( A_0 \)-eigenvalues \( \{e^{\pm j\theta_i T_s}\} \) are distinct and strictly complex, and \( v_i^H h \neq 0 \) for all \( i = 1, \ldots, d \). Then for all distinct pairs of points \( x, y \in M \), \( F \) satisfies (5) with constants \( C := d \left( \frac{\kappa_1^2}{A_1^2} + \frac{\kappa_2^2}{A_2^2} \right) \) and \( \delta := \delta_0 + \delta_1(M) \), where:

\[
\delta_0 := \frac{A_2\kappa_2^2 - A_1\kappa_1^2}{A_2\kappa_2^2 + A_1\kappa_1^2},
\]

\[
\delta_1(M) := \frac{(2d - 1)\nu}{M} \left( \frac{2A_2\kappa_2^2}{A_2\kappa_2^2 + A_1\kappa_1^2} \right).
\]

**Proof:** The proof of this theorem can be found in Appendix A.

We first note that the sufficient conditions of this theorem are the same as those in Theorem III.1, except that the required number of measurements is larger to ensure specific guarantees on the conditioning number \( \delta \). Also, note that this theorem requires an observation function with a particular norm \( \|h\|_2^2 = \frac{2d}{M} \). This normalization is to remove from \( C \) any dependence on the number of measurements \( M \) and the dimension of the attractor \( 2d \) (since \( \kappa_1^2 \) and \( \kappa_2^2 \) both scale inversely with \( d \)). The normalization plays no
other significant role in the proof (and therefore could be eliminated without losing generality, but at the expense of clarity).

To understand the implications of Theorem III.2, we examine the behavior of the conditioning number \( \delta \) as it is the main quantity of interest. In the theorem statement, \( \delta \) is a combination of \( \delta_0 \) (which does not depend on \( M \)) and \( \delta_1(M) \) which is positive for all \( M \) and for which \( \lim_{M \to \infty} \delta_1(M) = 0 \). Thus, we see that by taking more observations one could drive the conditioning guarantee for the mapping to \( \delta_0 \), but not below. In other words, some system and measurement pairs will have a plateau preventing the conditioning guarantee for the delay coordinate map from improving beyond a fundamental limit. This is in contrast with typical RIP results in CS where the conditioning can be continually improved by taking more measurements. Indeed, in order to get arbitrarily good conditioning we would need \( \delta_0 = 0 \), which happens if and only if

\[
A_2\kappa_2^2 - A_1\kappa_1^2 = 0 \iff \frac{A_2}{A_1} = \frac{\kappa_1^2}{\kappa_2^2} = 1.
\]

Recall that \( A_1 = A_2 \) implies that the attractor \( \mathcal{M} \) maximally fills the subspace spanned by \( V \) and \( \kappa_1 = \kappa_2 \) means that the observation function \( h \) projects equally onto the \( A \)-eigenvectors. Thus even with an infinite number of measurements, the delay coordinate map can only be guaranteed to be an isometry when the system and observation function maximally fill and measure the subspace containing the attractor.

The quantity \( \delta_1(M) \) can be used to determine the number of measurements necessary to ensure that the conditioning number \( \delta \) is within \( \epsilon \) of the optimal value \( \delta_0 \). To find the required number of measurements to meet this target \( \hat{M}(\epsilon) \), we set \( \delta_1(M) = \epsilon \) and solve (6) for \( M \) to get

\[
\hat{M}(\epsilon) = \frac{(2d - 1)\nu}{\epsilon} \left( \frac{2A_2\kappa_2^2}{A_2\kappa_2^2 + A_1\kappa_1^2} \right). \tag{7}
\]

By multiplying the numerator and denominator by \( \frac{1}{A_2\kappa_2^2} \) and noting that \( 0 < \frac{A_1\kappa_1^2}{A_2\kappa_2^2} \leq 1 \), we can deduce that \( \frac{(2d - 1)\nu}{\epsilon} \leq \hat{M}(\epsilon) < \frac{2(2d - 1)\nu}{\epsilon} \). While linear scaling with \( d \) is in line with state-of-the-art CS results, we see that in contrast to typical CS results \( \hat{M}(\epsilon) \) does not depend on the ambient dimension \( N \). Also note that \( \hat{M}(\epsilon) \) depends on \( \nu \), which is a worst case bound capturing the unfavorable joint relationships in the \( A \)-eigenvalues. Our subsequent simulation experiments will show the parameter \( \nu \) may be pessimistic in
some cases. While examining the proof has provided some insight into certain systems and observation functions where this quantity may be sharpened, we will not discuss those cases here for the sake of clarity and generality.

One immediate application of this theorem is that we can guarantee the A-RIP for the delay coordinate map and system pair with condition $\delta < 1$. This is made clear in the following corollary.

**Corollary III.1.** Suppose we have a linear system of class $\mathcal{A}(d)$, observation function $h$ and sampling time $T_s$ such that the conditions of Theorem III.2 are satisfied. Choose any $\epsilon > 0$. If the delay coordinate map $F$ defined in (4) has a number of delays $M$ chosen to satisfy $M > \hat{M}(\epsilon)$, with $\hat{M}(\epsilon)$ defined in (7), then the system and delay coordinate map satisfy the A-RIP with conditioning $\delta = \delta_0 + \epsilon$.

**Proof:** This corollary follows essentially immediately from Theorem III.2 and the calculation in (7).

After choosing $C$ as in Theorem III.2, we note that the condition $M > \hat{M}(\epsilon)$ is equivalent to $\delta_1(M) < \epsilon$. Thus, $\delta(M)$ defined in Theorem III.2 satisfies $\delta(M) = \delta_0 + \delta_1(M) < \delta_0 + \epsilon$, and Theorem III.2 then guarantees that the system and measurement pair satisfies A-RIP with conditioning $\delta + \epsilon$. 

---

**IV. Simulation experiments**

While the main result in Theorem III.2 is encouraging, it remains to be shown that the theoretical quantities actually reflect the salient embedding characteristics seen in system and measurement combinations. For example, it is important to know if the fundamental limits on the embedding quality $\delta(M)$ are artifacts of our proof technique or are empirically observed. If these limits on the embedding quality are actually present, it is also important to know if the related bounds are tight, both in their asymptotic values and in terms of their convergence speed as $M$ increases. This section will use a series of simple simulations to explore several aspects of the theoretical results expressed in Theorem III.2.

As a general approach, each simulation below will create an observation function $h$ and a test system of dimension $N = 50$ in class $\mathcal{A}(d)$ (defined by $\mathcal{A}$-eigenvalues and $\mathcal{A}$-eigenvectors) so that the conditions of Theorem III.2 are satisfied. We choose the arbitrary initial point $x_0$ defining the attractor such that $\alpha_0 = [1, \cdots, 1]^T$ and $x_0 = V \alpha_0$, and we assume a sample time of $T_s = 1$. For a single trial, we generate a random pair of points on the attractor $x$ and $y$ by choosing uniform random numbers $t_x, t_y$ from $(0, 10000)$ and assigning $x = V e^{A t_x} \alpha_0$ and $y = V e^{A t_y} \alpha_0$. In other words, the system is started from
the (arbitrary) initial condition and stopped after a random amount of time to get a single point on the attractor. We then vary $M$ from 1 to 200, and run 1000 trials for each $M$ (renormalizing $h$ for each $M$ as per Theorem III.2). For each trial we calculate the quality of the conditioning $Q(x, y) = \frac{\|F(x) - F(y)\|_2^2}{\|x - y\|_2^2}$, and for each $M$ record the largest and smallest value of $Q(x, y)$ (denoted $\max\{Q\}$ and $\min\{Q\}$, respectively) as a way to quantify the conditioning changes with the number of measurements. In the subsequent plots the dotted lines represent $C(1 \pm \delta_0)$, showing the theoretical asymptotic bounds on the conditioning quality $Q(x, y)$, and the dashed lines are the theoretical bounds on the conditioning $C(1 \pm \delta(M))$ given by Theorem III.2.

A. Bounds on the embedding quality

One of the fundamental characteristics of Theorem III.2 is that in general, the bound on the embedding quality $\delta(M)$ approaches $\delta_0 \neq 0$ as $M$ increases rather than approaching zero as is more typical in the analogous CS results. The first question to ask is whether pairs of systems and observation functions can actually display such a plateau as predicted, or whether the conditioning instead continually improves with more measurements. To demonstrate this effect, we generate a simulation as described above with $d = 3$, choosing the $A$-eigenvalues $\{\theta_i\}_{i=1}^d$ uniformly at random from $(0, \pi)$, and taking care to ensure that the resulting $A\Phi$-eigenvalues are distinct and strictly complex to satisfy the conditions of Theorem III.2. We then create the $A$-eigenvectors by letting $v_i = \frac{1}{\sqrt{2}}(e_{2i-1} + je_{2i})$, where $\{e_i\}$ are the canonical basis vectors in $\mathbb{R}^N$. This choice of $A$-eigenvectors ensures that $A_1 = A_2$. To generate a generic observation function $h$, we first create a vector $c \in \mathbb{R}^N$ such that $c = \sum_{i=1}^d ((1 + w_{2i-1}) \text{Re}\{v_i\} + (1 + w_{2i}) \text{Im}\{v_i\})$, where the $\{w_i\}$ are i.i.d. Gaussian random variables of zero mean and variance $\frac{1}{10}$. Thus $c$ is a (random) linear combination of the vectors that form the subspace of the attractor. For each $M$ we let $h = h(M) = \sqrt{\frac{2d}{M}} \frac{c}{\|c\|_2}$ so that $\|h\|_2^2 = \frac{2d}{M}$ to meet the conditions of Theorem III.2. Note that the small variance of $\{w_i\}$ produces $\{|v_i^H h|^2/\|h\|_2^2\}$ centered tightly around 1, making $\delta_0$ small (due to $A_1 = A_2$ and $\kappa_1, \kappa_2$ both close to 1).\footnote{The random variables $\{w_i\}$ are used to ensure that $\kappa_1, \kappa_2$ are close to, but not exactly equal to 1. The case where $\kappa_1 = \kappa_2 = 1$ is considered in the simulation in Figure 2(b).}

The specific parameters in this simulation are shown in Table I. The results for this simulation are shown in Figure 2(a). We see from the behavior of $\max\{Q\}$ and $\min\{Q\}$ that the embedding does indeed reach a fundamental limit where the conditioning does not
TABLE I
Parameters for the simulation shown in Figure 2(a). In this case the relevant quantities are $A_1 = A_2 = 1$, $\kappa_1 = 0.8346$, $\kappa_2 = 1.1637$, $\nu = 5.6954$ and $\delta_0 = 0.1647$.

<table>
<thead>
<tr>
<th>Index $i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_i$ (rad)</td>
<td>2.3129</td>
<td>0.1765</td>
<td>1.4861</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$|v_i^x/h_i^x|^2/|h_i|^2$</td>
<td>0.8346</td>
<td>1.1637</td>
<td>1.0017</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\lambda_i(V^H V)$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 2. Simulations exploring the asymptotic bounds on the conditioning of the delay coordinate map. Plotted are the largest and smallest value of $Q(x, y)$ (depicted by $\max\{Q\}$ and $\min\{Q\}$ respectively) attained by the 1000 pairs of $x, y$ for each $M$. The dotted (red) lines represent the values of $C(1 \pm \delta_0)$ and $C$, and the dashed (black) lines are the theoretical values of $C(1 \pm \delta(M))$. (a) In this simulation, $A_1 = A_2$ but $\kappa_1 \neq \kappa_2$, thus a plateau on the conditioning is seen. (b) In this simulation, $A_1 = A_2$ and $\kappa_1 = \kappa_2$. As expected, the conditioning number asymptotically reaches 0 as $M$ grows. (c) In this simulation, $A_1 \neq A_2$ and $\kappa_1 \neq \kappa_2$ and the predicted asymptotic values of the conditioning are not tight.

improve with more measurements. Furthermore, we see in this case that this plateau is correctly captured by the value $C(1 \pm \delta_0)$ as described in Theorem III.2. Additionally, the bounds $C(1 \pm \delta(M))$ do contain $\max\{Q\}$ and $\min\{Q\}$ as expected from the theorem, and the characteristic shape of these curves seems to qualitatively reflect the empirically observed convergence of the conditioning number.
As confirmation, we also verify the implication of Theorem III.2 that system and measurement combinations can be constructed where the conditioning can be made arbitrarily good with more measurements (akin to the more typical CS results). To show this, we create another system with the same $A$-eigenvalues and $A$-eigenvectors as in the previous simulation, with the latter implying that $A_1 = A_2$. For the observation function, we first define $c = V[1,\cdots,1]^T$, and for each $M$ we let $h = h(M) = \sqrt{\frac{2d}{M}} \frac{c}{\|c\|_2}$ as before. One can verify this choice results in $|v_i^H h|/\|h\|_2 = 1$ for all $i$, and thus $\kappa_1 = \kappa_2$. The parameters of this experiment are summarized in Table II.

<table>
<thead>
<tr>
<th>Index $i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_i$ (rad)</td>
<td>2.3129</td>
<td>0.1765</td>
<td>1.4861</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$</td>
<td>v_i^H h</td>
<td>/|h|_2$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>$\lambda_i(V^H V)$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**TABLE II**

Parameters for the simulation shown in Figure 2(b). The experiment was chosen such that $A_1 = A_2 = 1$ and $\kappa_1 = \kappa_2 = 1$, so that $\delta_0 = 0$. As the $A$-eigenvalues are the same as in the previous experiment, $\nu$ remains at 5.6954.

With this choice of parameters such that $A_1 = A_2$ and $\kappa_1 = \kappa_2$, Theorem III.2 indicates that $\delta_0 = 0$ so that $\lim_{M \to \infty} \delta(M) = 0$. Figure 2(b) shows the results of running the simulation in the same manner as before. The values of $\max\{Q\}$ and $\min\{Q\}$ clearly converge to $C$ as expected, showing that in this case the conditioning of the embedding can indeed be made arbitrarily good by taking more measurements.

Although Theorem III.2 indicates that a finite limit on the conditioning number is always reached when either $A_1 \neq A_2$ or $\kappa_1 \neq \kappa_2$, this bound is not always tight and the predicted plateau level of $C(1 \pm \delta_0)$ may be conservative. To show this, we construct a similar simulation as above, now setting the $A$-eigenvectors to be $v_i = \frac{1}{\sqrt{\|a_i\|^2 + \|b_i\|^2}}(a_i + jb_i)$, where $\{a_i, b_i\}$ are randomly constructed vectors in $\mathbb{R}^N$ whose entries are i.i.d. zero-mean Gaussian random variables with a variance of 1. We keep the $A$-eigenvalues the same and generate $h$ in the same manner as the first experiment shown in Figure 2(a).

The specific parameters for this experiment are shown in Table III, where we see that indeed $A_1 \neq A_2$ and $\kappa_1 \neq \kappa_2$.

Figure 2(c) shows the results of running the simulation in the same manner as before. We see that although a limit on the conditioning number is reached as predicted by Theorem III.2, the predicted plateau level of $C(1 \pm \delta_0)$ is not tight and the conditioning can be better than that predicted by $\delta_0$. 

18
TABLE III
Parameters for the simulation shown in Figure 2(c). We see that $A_1 = 0.4315$, $A_2 = 1.5316$, $\kappa_1 = 1.1318$ and $\kappa_2 = 1.8138$. Since the $A$-eigenvalues are the same as in the first simulation shown in Figure 2(a), $\nu$ remains the same at 5.6954. We also calculate $\delta_0 = 0.7010$.

| $i$ | $\theta_i$ (rad) | $|\nu^TH_i^*||H_i||$ | $\lambda_i(V^H V)$ |
|-----|------------------|----------------------|---------------------|
| 1   | 2.3129           | 1.8138               | 1.5316              |
| 2   | 0.1765           | 1.2064               | 1.3058              |
| 3   | 1.4861           | 1.1318               | 1.1294              |
| 4   | —                | —                    | 0.8372              |
| 5   | —                | —                    | 0.7644              |
| 6   | —                | —                    | 0.4315              |

Fig. 3. Examining the effect of the $A$-eigenvalues on the convergence speed of the conditioning. (a) In this simulation, $d = 1$ and we test $\theta = \frac{\pi}{200}, \frac{\pi}{100}$ and $\frac{\pi}{40}$. As expected, the closer $\theta$ is to $\frac{\pi}{2}$, the faster the rate of convergence of $\delta(M)$ to $\delta_0$. (b) In this simulation, $d = 3$ and we vary between 3 sets of $A$-eigenvalues with different values of $\nu$. As expected, the set of eigenvalues that gives the smallest $\nu$ provides the fastest rate of convergence of $\delta(M)$ to $\delta_0$ and vice versa.

B. Convergence Speed

In the simulations of the previous section we concentrated on the conditioning limits predicted by Theorem III.2, ignoring issues of the speed of convergence to those limits. Examining the formula for $\delta_1(M)$ in Theorem III.2, we see that the $A$-eigenvalues (via the parameter $\nu$) affect the convergence speed of $\delta(M)$ to its asymptotic value of $\delta_0$. In particular, the convergence speed scales with $1/\nu$, which is also demonstrated in (7) where the number of measurements $\hat{M}(\epsilon)$ necessary to get the conditioning $\delta$ within $\epsilon$ of the best possible value ($\delta_0$) is proportional to $\nu$.

For ease of analysis, we first consider the case where $d = 1$, meaning that $\nu = |\sin(\theta)|^{-1}$ (since $T_s = 1$) where $\pm j\theta$ are the sole $A$-eigenvalues. In this case, $|\sin(\theta)|^{-1} \geq 1$ with the minimum attained when $\theta = \frac{\pi}{2} + k\pi$ for any integer $k$. The closer $\theta$ is to $\frac{\pi}{2} + k\pi$, the faster the convergence of $\delta(M)$ to $\delta_0$. This is illustrated by the following simulation where the $A$-eigenvectors are chosen such that $A_1 = A_2$, $A_2$,
and the observation function is chosen randomly as in the experiment shown in Figure 2(a) (except with $d = 1$). Figure 3(a) plots $\max\{Q\}$ and $\min\{Q\}$ for $\theta = \frac{\pi}{200}, \frac{\pi}{100}$, and $\frac{\pi}{40}$, showing that Theorem III.2 correctly captures that the convergence speed to the asymptotic value of $C(1 \pm \delta_0)$ varies inversely with the value of $\theta$.

When $d > 1$, the joint relationship of the $A$-eigenvalues (not just their individual values) determines $\nu$, and subsequently the convergence speed. One can see intuitively in the definition of $\nu$ that $A$-eigenvalues which are maximally spread out should produce favorable convergence speeds. To illustrate this, we generate a simulated system with $d = 3$, choosing the $A$-eigenvectors such that $A_1 = A_2$, and generating an observation function $h$ randomly (as in the experiment in Figure 2(a)). We also choose three sets of $A$-eigenvalues: two uniformly random sets, and one set that are slight perturbations of equally spaced points around the unit circle according to $\theta_p = \frac{p\pi}{d+1}$ (the choices of $\theta_p$ and their respective $\nu$ are given in Table IV). Figure 3(b) shows the results of the simulation, with the $\max\{Q\}$ and $\min\{Q\}$ curves showing clearly that $\nu$ indeed controls the speed of convergence of $\delta(M)$ as predicted.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>0.7836</td>
<td>1.5864</td>
<td>2.3566</td>
<td>2.6619</td>
</tr>
<tr>
<td>Set 2</td>
<td>0.0491</td>
<td>1.5737</td>
<td>2.3490</td>
<td>20.3851</td>
</tr>
<tr>
<td>Set 3</td>
<td>0.0212</td>
<td>1.5684</td>
<td>2.3549</td>
<td>47.1388</td>
</tr>
</tbody>
</table>

TABLE IV
Choice of $\{\theta_i\}$ (in radians) for the experiment in Figure 3(b) and their respective $\nu$ value.

Given that Theorem III.2 seems to be correctly capturing the convergence speed dependence on $\nu$, the last facet of the problem to explore is the tightness of this bound. Specifically, given a system of class $\mathcal{A}(d)$ and an observation function $h$, it is often of interest to estimate the minimum number of measurements $\left(\widehat{M}(\epsilon)\right)$ needed to ensure that for any $M' \geq M$ the conditioning number $\delta(M')$ is at most $\epsilon$ above the asymptotic level of $\delta_0$ (such an estimate is given in (7)). To examine this, we refer back to the simulation shown in Figure 2(a) with parameters given in Table I. Fixing $\epsilon = 0.2$, Figure 4(a) re-plots $\max\{Q\}$ together with the line $C(1 + \delta_0 + \epsilon)$. Using the given parameters and (7) we calculate that $\widehat{M}(\epsilon) \approx 166$. Note that this value is also the intersection of the curve $C(1 + \delta(M))$ with the

11 The slight perturbation is used for plotting convenience so all three curves converge to the same asymptotic value. If exactly equally spaced eigenvalues are used, the attractor is sampled uniformly and the convergent value will be inside $C(1 \pm \delta_0)$, making comparative plots difficult.
line $C(1 + \delta_0 + \epsilon)$. Figure 4(a) shows that $\max\{Q\}$ actually met this tolerance with only around 30 measurements. Thus, although the theoretical value of $\widehat{M}(\epsilon)$ given by (7) is correct, it is pessimistic in at least this particular case.

To demonstrate that the linear dependence of $\widehat{M}(\epsilon)$ on $\nu$ is correctly captured in the theorem, we restrict ourselves to $d = 1$. Recall that when $d = 1$, $\nu = |\sin(\theta)|^{-1}$ (since $T_s = 1$) where $\pm j\theta$ are the sole $\mathcal{A}$-eigenvalues. We repeat the simulation shown in Figure 3(a), this time using 100 values of $\theta$ equally spaced between $(0, \pi/2)$. Fixing $\epsilon = 0.1$, for each value of $\theta$ we note the value of $M$ where for all $M' > M$, $\max\left\{\frac{\max(Q)}{C} - 1, 1 - \frac{\min(Q)}{C}\right\} < \delta_0 + \epsilon$. We call this value the “actual” $\widehat{M}(\epsilon)$, in contrast to the “theoretical” $\widehat{M}(\epsilon)$ given by (7). Figure 4(b) shows these actual and theoretical values of $\widehat{M}(\epsilon)$ as a function of $\theta$. This comparison shows that while the theoretical $\widehat{M}(\epsilon)$ captures the same trend as the actual $\widehat{M}(\epsilon)$, the theoretical estimate can be pessimistic compared to the empirical values (though it is not clear if the theoretical bounds are achieved by some systems).

V. CONCLUSION

The main result of this paper has established that a delay coordinate map (using linear observation functions) can form a stable embedding for all pairs of points on the attractor of a linear dynamical system of class $\mathcal{A}(d)$. The explicit, deterministic and non-asymptotic sufficient conditions we give for this stable
embedding yield several observations about the embedding itself and favorable properties of system and measurement pairs. For example, the minimum number of delays \( M \) of the delay coordinate map scales linearly with the attractor dimension but is independent of the system dimension. This is in contrast with typical CS results, where the number of compressive measurements also scales logarithmically with the system dimension (but interestingly does parallel recent improvements in these bounds for the stable embedding of manifolds [17]). Additionally, for many system and measurement pairs, the conditioning number \( \delta(M) \) reaches a non-zero asymptotic value of \( \delta_0 \) with increasing \( M \). This “plateau effect” is again in contrast with typical CS results where the conditioning of the stable embedding can be continually improved by increasing the number of measurements. Furthermore, the convergence speed of the embedding quality to this limit is governed by the joint relationship of the system eigenvalues, which capture the relative speed with which the system explores the different dimensions of the state space (i.e., more diversity in these speeds results in faster convergence).

While the comparisons with standard CS results reveal these interesting and non-intuitive technical differences between the results in each case, these discrepancies actually point to a much deeper difference in the problem setups that must be appreciated when embedding attractors of dynamical systems. Perhaps the easiest way to see this is to consider that in the present case of delay coordinate maps, while the number of measurements doesn’t scale with the ambient system dimension, the total number of measurements may in fact have to be larger than the system dimension \( (M > N) \) in order to make a particular conditioning guarantee. In the typical CS case, this would of course be a ridiculous proposition. If the RIP property required \( (M > N) \) random measurements (e.g., due to very large constants in the typical sufficient conditions), one would likely abandon the CS strategy and simply take \( N \) uncoded measurements (e.g., in the canonical basis). However, in the case of delay coordinate maps for dynamical systems, this luxury is simply not available. For example, the observers often do not have any control over the choice of observation function \( h \), and in these cases cannot simply change the way the system is measured. But, more importantly, even if we were given complete control over \( h \), it is only a “seed” that is used in producing the whole measurement process. One can view the entire set of measurements as being generated by repeatedly forcing this observation function through the dynamics of the system (seen explicitly in writing the delay coordinate map in (4)). Said another way, because there is only a single observation function for the system, the total measurement process for a delay coordinate map is
beholden to the dynamics of the system itself to provide sufficient diversity to make the measurements informative. Therefore, even with complete control over the observation function, delay coordinate maps represent a highly restricted total measurement process that cannot be completely controlled (without access to and control over the system that is hidden and in need of measurement).

Characterizing the delay coordinate map embeddings for attractors of linear dynamical systems with linear observation functions is a subset of the more general problem of characterizing these embeddings for attractors of nonlinear systems and general observation functions. From the results here, we conclude that there is reason to be optimistic that similar stability results can be obtained for this more general case of interest. Furthermore, these results also lead us to conclude that there are several issues that differ from standard CS results and will need to be carefully considered in any generalization. Finally, even in the case of linear systems, there are several areas of future work that would be beneficial. In particular, the parameter $\nu$ is a worst case bound characterizing the relationships between $A$-eigenvalues (determining the relative speeds on different subspaces of the attractor). In the current definition, $\nu$ can be completely dominated by one unfavorable pairing of these $A$-eigenvalues, resulting in bounds that can be pessimistic for many systems. An improvement to this approach that admits a clean statement of the convergence speed but is refined to capture subtleties of the system parameters would be valuable.

**APPENDIX A**

**Proof of Stable Takens’ Embedding Theorem**

Because Theorems III.1 and III.2 are very similar in structure, we will essentially lay out the proof approach for both of them together in this section and then separately establish the necessary details for each result. Before proceeding with the specific proofs, we will introduce some notation and preliminary results that will be useful.

**A. Notation and preliminaries**

1) **Frame theory:** Drawing on some terminology from the field of frame theory, we say that a sequence of vectors $\{g_i\}_{i=1}^M$ in $\mathbb{C}^K$, $M \geq K$, forms a frame [25] for $\mathbb{C}^K$ if there exists two real constants $0 < B_1 \leq B_2 < \infty$ such that for all $\alpha \in \mathbb{C}^K$,

$$B_1 \|\alpha\|^2_2 \leq \sum_{i=1}^M |\langle g_i, \alpha \rangle|^2 = \|G\alpha\|^2_2 \leq B_2 \|\alpha\|^2_2,$$

where $G = \{g_i\}_{i=1}^M$. In this context, $\alpha$ is the vector representing the state of the system, $G\alpha$ is the measurement vector, and $B_1$ and $B_2$ are the frame bounds.
where \( G^H = (g_1 | g_2 | \cdots | g_M) \in \mathbb{C}^{K \times M} \), the concatenation of the \( \{g_i\}_{i=1}^M \), is called the frame analysis operator and \( B_1, B_2 \) are called the frame bounds. The frame bounds can be defined as \( B_1 = \lambda_{\text{min}} \) and \( B_2 = \lambda_{\text{max}} \), where \( \lambda_{\text{min}} \) and \( \lambda_{\text{max}} \) are the minimum and maximum eigenvalues of \( G^H G \in \mathbb{C}^{K \times K} \).

2) Linear delay coordinate maps: Because the attractor \( \mathcal{M} \) is contained in the span of the columns of \( V \), for any \( x, y \in \mathcal{M} \) we can write \( x = V \alpha_x \) and \( y = V \alpha_y \) for some complex coefficients \( \alpha_x, \alpha_y \in \mathbb{C}^{2d} \). Using \( F \) to denote the delay coordinate map for a linear system with flow matrix \( \Phi \) and observation function \( h \) as described in (4), the \( k \)-th row (for \( k = 1, \cdots, M \)) of the vector \( F(x) - F(y) \) can be written:

\[
\begin{align*}
  h^T \left( \Phi^{k-1}(x - y) \right) &= h^T \left( \Phi^{k-1}V(\alpha_x - \alpha_y) \right) \\
  &= h^T \left( VD^{k-1}(\alpha_x - \alpha_y) \right) = (g_k, \alpha_x - \alpha_y),
\end{align*}
\]

where

\[
g_k = \left( h^T V D^{k-1} \right)^H = \left[ (v_1^H h)e^{-j(k-1)\theta_s T_s}, (v_1^H h)e^{j(k-1)\theta_s T_s}, \cdots, (v_d^T h)e^{-j(k-1)\theta_s T_s}, (v_d^H h)e^{j(k-1)\theta_s T_s} \right]^H,
\]

and \( D \) is the diagonal matrix comprised of \( \Phi \)-eigenvalues as defined in Section II-A. Thus, we have:

\[
\|F(x) - F(y)\|_2 = \sum_{k=1}^M \| (g_k, (\alpha_x - \alpha_y)) \|^2 = \|G(\alpha_x - \alpha_y)\|_2^2,
\]

where \( G \in \mathbb{C}^{M \times 2d} \) is the concatenation of \( \{g_k\} \) as described above. In this following, \( G \) is fixed to be this matrix given here.

3) Eigenvalue bounds: It will be important in the following proofs to determine bounds on the extreme eigenvalues of the matrix \( G^H G \). To that end, we first introduce the well-known Gresgorin Circle Theorem, which we state here for notational convenience:

**Theorem A.1** (Gresgorin Circle Theorem [20]). The eigenvalues of a \( K \times K \) matrix \( A \) all lie in the union of the Gresgorin disks of \( A \). The Gresgorin disk \( \mathcal{D}_i \) for \( i = 1, \cdots, K \), is defined as \( \mathcal{D}_i = \{ x \in \mathbb{C} : |x - C_i| \leq \tilde{r}_i \} \), where \( \tilde{r}_i := \sum_{j=1}^K |(A)_{i,j}| \) is the radius, and \( C_i := (A)_{i,i} \) is the center of the \( i \)-th Gresgorin disk. Thus \( \lambda(A) \subset \bigcup_{i=1}^K \mathcal{D}_i \), where \( \lambda(A) = \{\lambda_1, \cdots, \lambda_K\} \), and \( \{\lambda_i\} \) are the eigenvalues of \( A \).

To apply the Gresgorin Circle Theorem in order to obtain the extrema eigenvalues of \( G^H G \), we introduce the following useful lemma that gives values for centers \( C_i \) and radii \( \tilde{r}_i \) of the Gresgorin disks
$D_i$ of $G^HG$.

**Lemma A.1.** For $i = 1, \cdots, d$, the centers and radii of the Gresgorin disks of $G^HG \in \mathbb{C}^{2d \times 2d}$ are:

$$C_{2i-1} = C_{2i} = |v_i^H h|^2 M$$

$$\tilde{r}_{2i-1} = \tilde{r}_{2i} = |v_i^H h|^2 \left( \sin(M \theta_i T_s) \right) + \sum_{p=1}^{d} |v_{p}^H h|^2 \left( \sin \left( \frac{(\theta_p - \theta_i) T_s}{2} \right) \right)$$

$$+ \sum_{p=1, p \neq i}^{d} |v_{p}^H h|^2 \left( \sin \left( \frac{(\theta_p + \theta_i) T_s}{2} \right) \right)$$

**Proof:** We can write $G^HG$ as

$$G^HG = \begin{pmatrix} g_1 & \cdots & g_M \end{pmatrix} \begin{pmatrix} g_1^H \\ \vdots \\ g_M^H \end{pmatrix} = \sum_{k=1}^{M} g_k g_k^H,$$

where we recall that $g_k$ is defined as in (8). Thus the $(p, q)$ entry of $G^HG$ can be expressed as: $(G^HG)_{p,q} = \sum_{k=1}^{M} g_k(p) g_k(q)^*$, where $g_k(p)$ denotes the $p$-th entry of the vector $g_k$. As such, the formation of $G^HG$ involves the calculation of sum of complex trigonometric polynomials due to the complex exponentials $(\{ e^{\pm j(k-1)\theta_p T_s} \})$ appearing in the terms of each $g_k$. A few separate cases need to be considered because of the differences in the even $(2p)$ and odd $(2p-1)$ numbered rows of $G^HG$ for all $p$. We first consider the even numbered rows. The diagonal terms actually have a fairly simple form:

$$(G^HG)_{2p,2p} = \sum_{k=1}^{M} g_k(2p) g_k(2p)^* = \sum_{k=1}^{M} |v_{p}^H h|^2 = M |v_{p}^H h|^2.$$

The adjacent term to the left is given by:

$$(G^HG)_{2p,2p-1} = \sum_{k=0}^{M-1} \left( (v_p^T h) e^{-j k \theta_p T_s} \right)^2 = (v_p^T h)^2 \sum_{k=0}^{M-1} \left( e^{-j 2 \theta_p T_s} \right)^k = (v_p^T h)^2 \sin(M \theta_p T_s) e^{-j (M-1) \theta_p T_s},$$

where the last expression follows from the standard formula for a finite geometric sum, pulling out common exponential factors, and using Euler’s formula. The other cross terms for all $p, q \in \{1, \ldots, d\}$
such that \( p \neq q \) can be derived similarly as:

\[
(G^H G)_{2p, 2q} = (v_p^THh)(v_q^Hh) \sum_{k=0}^{M-1} \left( e^{-j2\left(\frac{\theta_p+\theta_q}{2}\right)T_s} \right)^k = (v_p^THh)(v_q^Hh) \frac{\sin(M\left(\frac{\theta_p-\theta_q}{2}\right)T_s)}{\sin(\left(\frac{\theta_p+\theta_q}{2}\right)T_s)} e^{-j(M-1)\left(\frac{\theta_p-\theta_q}{2}\right)T_s},
\]

\[
(G^H G)_{2p, 2q-1} = (v_p^THh)(v_q^Hh) \sum_{k=0}^{M-1} \left( e^{-j2\left(\frac{\theta_p-\theta_q}{2}\right)T_s} \right)^k = (v_p^THh)(v_q^Hh) \frac{\sin(M\left(\frac{\theta_p+\theta_q}{2}\right)T_s)}{\sin(\left(\frac{\theta_p+\theta_q}{2}\right)T_s)} e^{-j(M-1)\left(\frac{\theta_p-\theta_q}{2}\right)T_s}.
\]

The relevant quantities for the odd numbered rows are given similarly as

\[
(G^H G)_{2p-1, 2p-1} = (G^H G)_{2p, 2p} = M|v_p^Hh|^2,
\]

\[
(G^H G)_{2p-1, 2p} = (G^H G)_{2p, 2p-1} = (v_p^Hh)^2 \frac{\sin(M\theta_pT_s)}{\sin(\theta_pT_s)} e^{j(M-1)\theta_pT_s},
\]

\[
(G^H G)_{2p-1, 2q} = (G^H G)_{2q, 2p-1} = (v_q^Hh)(v_p^Hh) \frac{\sin(M\left(\frac{\theta_p+\theta_q}{2}\right)T_s)}{\sin(\left(\frac{\theta_p+\theta_q}{2}\right)T_s)} e^{j(M-1)\left(\frac{\theta_p+\theta_q}{2}\right)T_s},
\]

\[
(G^H G)_{2p-1, 2q-1} = (v_p^Hh)(v_q^Hh) \frac{\sin(M\left(\frac{\theta_p-\theta_q}{2}\right)T_s)}{\sin(\left(\frac{\theta_p-\theta_q}{2}\right)T_s)} e^{j(M-1)\left(\frac{2\theta_p-\theta_q}{2}\right)T_s}.
\]

Finally we note that many of the above complex quantities only differ in their phase because of symmetry in the summations, making their magnitudes equal when calculating the radii of the Gresgorin disks. The expressions for \( C_i \) and \( \tilde{r}_i \) in the lemma are obtained simply by applying the notation of the Gresgorin Circle Theorem to the calculated magnitudes of the entries of \( G^H G \).

\[\blacksquare\]

**B. General proof approach**

Using the preliminaries above, we can now sketch out the general approach for the proof of both theorems below. Essentially, the theorems result from using (or establishing) the following three facts:

1) If \( G^H G \in \mathbb{C}^{2d \times 2d} \) is established to be full rank, then \( \{g_k\}_{k=1}^M \) form a frame in \( \mathbb{C}^{2d} \). Thus there exists \( 0 < B_1 \leq B_2 < \infty \) such that

\[
B_1 \|\alpha_x - \alpha_y\|^2_2 \leq \|F(x) - F(y)\|^2_2 \leq B_2 \|\alpha_x - \alpha_y\|^2_2,
\]

holds for all distinct pairs of points \( x, y \in \mathcal{M} \). In particular, to establish conditioning guarantees, we can let \( B_1 \) and \( B_2 \) be the smallest and largest eigenvalues of \( G^H G \) (respectively) and determine bounds on those important quantities.
2) Next, we use the fact that $\|x - y\|_2^2 = (\alpha_x - \alpha_y)^H V^H V (\alpha_x - \alpha_y)$ to get:

$$A_1 \|\alpha_x - \alpha_y\|^2_2 \leq \|x - y\|^2_2 \leq A_2 \|\alpha_x - \alpha_y\|^2_2,$$

(10)

where $A_1$ and $A_2$ are the smallest and largest eigenvalues of $V^H V \in \mathbb{C}^{2d \times 2d}$, respectively. By the definition of $V$ we know that $V^H V$ is well-defined and full rank, meaning that $0 < A_1 \leq A_2 < \infty$.

3) Putting (9) and (10) together, we get: $0 < \frac{B_1}{A_2} \leq \frac{\|F(x) - F(y)\|^2_2}{\|x - y\|^2_2} \leq \frac{B_2}{A_1} < \infty$, where the bounds $\frac{B_1}{A_2}$ and $\frac{B_2}{A_1}$ can be manipulated to get the A-RIP (2). Specifically, we can set the A-RIP constants to be $C = \frac{1}{2} \left( \frac{B_1}{A_2} + \frac{B_2}{A_1} \right)$ and $\delta = 1 - \frac{B_1}{CA_2}$.

C. Proof of Theorem III.1

Proof: For Theorem III.1, we follow the three steps detailed in Appendix A-B, where we only need to show that $G^H G$ is indeed full rank given the conditions of the theorem. Consider first the case when $M = 2d$, where showing $G^H G$ is full rank is equivalent to showing $\text{det}(G^H G) = \text{det}(G)^2 > 0$. The matrix $G$ can be expressed in terms of a product of a Vandermonde matrix and a diagonal matrix:

$$G = \begin{pmatrix}
(v_1^T h) & (v_1^H h) & \cdots & (v_d^T h) & (v_d^H h) \\
e^{-j\theta_1 T_s} (v_1^T h) & e^{j\theta_1 T_s} (v_1^H h) & \cdots & e^{-j\theta_d T_s} (v_d^T h) & e^{j\theta_d T_s} (v_d^H h) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
e^{-j2\theta_1 T_s} (v_1^T h) & e^{j2\theta_1 T_s} (v_1^H h) & \cdots & e^{-j2\theta_d T_s} (v_d^T h) & e^{j2\theta_d T_s} (v_d^H h)
\end{pmatrix} = \begin{pmatrix}
1 & 1 & \cdots & 1 & 1 \\
e^{-j\theta_1 T_s} & e^{j\theta_1 T_s} & \cdots & e^{-j\theta_d T_s} & e^{j\theta_d T_s} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
e^{-j2\theta_1 T_s} & e^{j2\theta_1 T_s} & \cdots & e^{-j2\theta_d T_s} & e^{j2\theta_d T_s}
\end{pmatrix} \begin{pmatrix}
v_1^T h \\
v_1^H h \\
\vdots \\
v_d^T h \\
v_d^H h
\end{pmatrix} = \tilde{M}^T \tilde{H},$$

where $\tilde{M}$ is the Vandermonde matrix with the $A_\theta$-eigenvalues as its parameters and $\tilde{H}$ is a diagonal matrix whose diagonal elements are made up of the projection of $h$ onto the $A$-eigenvectors. Thus, $\text{det}(G) = \text{det}(\tilde{M}) \text{det}(\tilde{H})$. One of the conditions of Theorem III.1 ensures that the $\{e^{\pm j\theta_i T_s}\}_{i=1}^d$ are distinct, which implies that the determinant of this square Vandermonde matrix [26, Ch 0] obeys $|\text{det}(\tilde{M})| > 0$. Also since $v_i^T h \neq 0$ for all $i = 1, \cdots, d$, we also know that $|\text{det}(\tilde{H})| > 0$. Therefore for
\[ M = 2d, \ \text{rank}(G^H G) = 2d. \] Since adding vectors to a frame does not change the rank of \( G^H G \) (i.e., frame bounds cannot be lowered by adding more vectors to the frame), it follows that if \( M \geq 2d \) then \( \text{rank}(G^H G) = 2d \) and the proof of Theorem III.1 is complete.

D. Proof of Theorem III.2

Proof: To prove Theorem III.2, we again follow the three steps detailed in Appendix A-B, this time establishing specific guarantees on the frame bounds \( B_1(M) \) and \( B_2(M) \) appearing in the first step. From Lemma A.1, we first observe that for all \( i \) we can bound the Gresgorin disk radii by

\[
\tilde{r}_{2i-1} = \tilde{r}_{2i} \leq \left( |v_i^H h|^2 + \sum_{p=1}^{d} |v_i^H h||v_p^H h| + \sum_{p=1}^{d} |v_i^H h||v_p^H h| \right) \nu \leq \left( 2d - 1 \right) \nu \kappa_2^2 ||h||^2 \nu.
\]

Noting that \( ||h||^2 = \frac{2d}{M} \), we see that for each \( i \), the Gresgorin disks of \( G^H G \) satisfy:

\[
D_{2i-1} = D_{2i} \subset [ |v_i^H h|^2 M - ||h||^2 (2d - 1) \nu \kappa_2^2, |v_i^H h|^2 M + ||h||^2 (2d - 1) \nu \kappa_2^2 ].
\]

Then applying the Gresgorin Circle Theorem, we get:

\[
\lambda(G^H G) \subset \bigcup_j D_j \subset \left[ 2d \kappa_1^2 - \frac{2d}{M} (2d - 1) \nu \kappa_2^2, 2d \kappa_2^2 + \frac{2d}{M} (2d - 1) \nu \kappa_2^2 \right].
\]

By choosing \( B_1(M) = 2d \left( \kappa_1^2 - \frac{(2d - 1)\nu \kappa_2^2}{M} \right) \) and \( B_2(M) = 2d \left( \kappa_2^2 + \frac{(2d - 1)\nu \kappa_2^2}{M} \right) \), and applying (10), we arrive at:

\[
\frac{B_1(M)}{A_2} \leq \frac{\|F(x) - F(y)\|_2^2}{\|x - y\|_2^2} \leq \frac{B_2(M)}{A_1} \tag{11}
\]

for all distinct pairs of points \( x, y \in \mathcal{M} \) and for all \( M \).

Now as \( M \to \infty \), \( B_1(M) \to 2d \kappa_1^2 \) and \( B_2(M) \to 2d \kappa_2^2 \). Thus in the limit of large \( M \), the lower and upper bounds of the inequality (11) approaches \( \frac{2d \kappa_1^2}{A_2} \) and \( \frac{2d \kappa_2^2}{A_1} \), respectively. We define the scaling constant \( C \) as the average of the asymptotic values of these lower and upper bounds: \( C := \frac{2d}{2} \left( \frac{\kappa_1^2}{A_2} + \frac{\kappa_2^2}{A_1} \right) \). Also define the conditioning number \( \delta(M) \) for a given \( M \) as the maximum deviation of the lower and upper bounds of (11) from \( C \), normalized by \( C \): \( \delta(M) := \max \left\{ 1 - \frac{B_1(M)}{CA_2}, \frac{B_2(M)}{CA_1} - 1 \right\} \). Looking at each of
the terms:

\[
1 - \frac{B_1(M)}{CA_2} = 1 - \frac{2d \left( \frac{\kappa_1^2 - (2d-1)\nu \kappa_2^2}{M} \right)}{\frac{2d}{2} \left( \kappa_1^2 + \kappa_2^2 \frac{A_2}{A_1^2} \right)} = \frac{A_2 \kappa_2^2 - A_1 \kappa_1^2 + \frac{2A_1(2d-1)\nu \kappa_2^2}{M}}{A_2 \kappa_2^2 + A_1 \kappa_1^2},
\]

\[
\frac{B_2(M)}{CA_1} - 1 = \frac{2d \left( \frac{\kappa_2^2 + (2d-1)\nu \kappa_1^2}{M} \right)}{\frac{2d}{2} \left( \frac{A_2 \kappa_2^2}{A_1^2} + \kappa_2^2 \right)} - 1 = \frac{A_2 \kappa_2^2 - A_1 \kappa_1^2 + \frac{2A_2(2d-1)\nu \kappa_1^2}{M}}{A_2 \kappa_2^2 + A_1 \kappa_1^2},
\]

we have (since \(A_1 \leq A_2\)) that \(\delta(M) = \frac{B_2(M)}{CA_1} - 1 = \frac{A_2 \kappa_2^2 - A_1 \kappa_1^2 + \frac{2A_2(2d-1)\nu \kappa_1^2}{M}}{A_2 \kappa_2^2 + A_1 \kappa_1^2}.\) We can then define \(\delta_0\) and \(\delta_1(M)\) as the first and second term of the sum above. Notice that \(\delta(M)\) represents a worst case bound on the deviation from \(C\), as we maximized over upper and lower bounds that may not be the same magnitude (i.e., in general \(C(1 - \delta(M)) \neq \frac{B_1(M)}{A_1}\)).

Finally, we recall that for the embedding conditioning number to be valid, we must have \(0 \leq \delta(M) < 1\). The first condition \(\delta(M) \geq 0\) is achieved by construction. The upper bound is equivalent to:

\[
\delta(M) < 1 \iff \frac{2A_2(2d-1)\nu \kappa_2^2}{M} < 2A_1 \kappa_1^2 \iff \frac{(2d - 1)\nu}{M} < \frac{A_1 \kappa_1^2}{A_2 \kappa_2^2},
\]

which is the condition for \(M\) required by the theorem statement, thus completing the proof.

\section*{References}


