

Supplementary Material: A Geometry-Aware Metric for Mode Collapse in Multivariate Time Series Generative Models

893 **A Dynamical Mode Decomposition: Details and Proofs**

894 **A.1 The Link Between the DMD operators in Continuous and Discrete Cases**

Given a dynamical system $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t; \mu)$, we linearly approximate the dynamics using DMD using the operator $\mathcal{A} \in \mathbb{R}^{n \times n}$, i.e.

$$\forall t, \quad \dot{\mathbf{x}}(t) = \mathcal{A}\mathbf{x}.$$

895 Discretizing time into intervals of Δt and capturing snapshots accordingly, we establish the relation-
896 ship between consecutive time steps in the following equation:

$$\forall k, \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \mathcal{A}\mathbf{x}_k\Delta t = (I + \Delta t\mathcal{A})\mathbf{x}_k. \quad (4)$$

897 For a time-step Δt that is sufficiently small, we can employ the first-order Taylor expansion of the
898 matrix $\exp(\Delta t\mathcal{A})$, expressed as:

$$\exp(\Delta t\mathcal{A}) \approx I + \Delta t\mathcal{A} \quad (5)$$

Therefore, from Equations 4 and 5, we conclude that:

$$\forall k, \quad \mathbf{x}_{k+1} \approx \exp(\Delta t\mathcal{A})\mathbf{x}_k.$$

Thus,

$$\mathbf{A}^* \approx \exp(\Delta t\mathcal{A}).$$

899 **A.2 Feasible Spectral Decomposition of the DMD Operator using Dimensionality Reduction**

900 Algorithm 2 presents the steps to compute the eigenvectors and eigenvalues of the DMD operator \mathbf{A}^*
901 using Singular Value Decomposition (SVD) for dimensionality reduction.

902 **A.3 DMD expansion**

903 We will proof the closed formula $\forall k, \quad \mathbf{x}_k = \sum_{j=1}^r \phi_j \lambda_j^k b_j = \Phi \Lambda^k \mathbf{b}$, using recursion.

904 For $k = 0$, we have,

$$\begin{aligned} \mathbf{x}_0 &= \mathbf{I}\mathbf{x}_0 \\ &= \Phi \Phi^\dagger \mathbf{x}_0 \\ &= \Phi \mathbf{b} \\ &= \Phi \Lambda^0 \mathbf{b} \end{aligned}$$

905 Let's now consider the equation hold for $k = 0, \dots, m$, we have,

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}^* \mathbf{x}_k \\ &= \mathbf{A}^* \Phi \Lambda^k \mathbf{b} \\ &= \Phi \Lambda \Lambda^k \mathbf{b} \\ &= \Phi \Lambda^{k+1} \mathbf{b}. \end{aligned}$$

906 Therefore, the equality holds for all $k \in \mathbb{N}$.

Algorithm 2 Dynamic Mode Decomposition

1. From collected snapshots of the system, build a pair of data matrices $(\mathbf{X}, \mathbf{X}')$.

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_0 & \mathbf{x}_1 & \cdots & \mathbf{x}_{m-1} \\ | & | & & | \end{bmatrix}, \mathbf{X}' = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ | & | & & | \end{bmatrix}$$

The closed formula of optimal DMD operator is

$$\mathbf{A}^* = \mathbf{X}'\mathbf{X}^\dagger$$

2. Compute the compact singular value decomposition (SVD) of \mathbf{X} :

$$\mathbf{X} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$$

where $\mathbf{U} \in \mathbb{C}^{n \times r}$, $\mathbf{\Sigma} \in \mathbb{C}^{r \times r}$, $\mathbf{V} \in \mathbb{C}^{m \times r}$ and $r \leq \min(m, n)$ is the rank of \mathbf{X} . Therefore,

$$\mathbf{A}^* = \mathbf{X}'\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^\dagger$$

3. Define a matrix

$$\tilde{\mathbf{A}} = \mathbf{U}^\dagger \mathbf{A}^* \mathbf{U} = \mathbf{U}^\dagger \mathbf{X}' \mathbf{V} \mathbf{\Sigma}^{-1},$$

since \mathbf{U} is a unitary matrix.

$\tilde{\mathbf{A}} \in \mathbb{R}^{r \times r}$ defines a low-dimensional linear model of the dynamical system on proper orthogonal decomposition (POD) coordinates.

4. Compute the eigen-decomposition of $\tilde{\mathbf{A}}$:

$$\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda},$$

where columns of $\mathbf{W} \in \mathbb{R}^{r \times r}$ are eigenvectors and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_r) \in \mathbb{R}^{r \times r}$ is a diagonal matrix containing the corresponding eigenvalues.

5. Return DMD modes Φ :

$$\Phi = \mathbf{X}'\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}.$$

Each column of Φ is an eigenvector of \mathbf{A} meaning a DMD mode ϕ_k corresponding to eigenvalue λ_k

907 B Proof of Theorem 3.6 - DMD Mode Geodesic

908 Theorem 3.6 [DMD Mode Geodesic]

909 Let $\mathcal{M}_k(\mathbf{X}), \mathcal{M}_k(\tilde{\mathbf{X}}) \in \mathbb{R}^{n \times k}$ be matrices whose columns form orthonormal bases of two k -
 910 dimensional subspaces of \mathbb{R}^n . Let $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_k)$ be the diagonal matrix of principal
 911 angles between the subspaces spanned by $\mathcal{M}_k(\mathbf{X})$ and $\mathcal{M}_k(\tilde{\mathbf{X}})$. Let $\Delta \in \mathbb{R}^{n \times k}$ be an orthonormal
 912 matrix such that

$$\mathcal{M}_k(\tilde{\mathbf{X}}) = \mathcal{M}_k(\mathbf{X}) \cos(\Theta) + \Delta \sin(\Theta). \quad (6)$$

913 Then, the geodesic linking $\mathcal{M}_k(\mathbf{X})$ and $\mathcal{M}_k(\tilde{\mathbf{X}})$ on the Grassmann manifold $\text{Gr}(k, n)$ is given by

$$\gamma(t) = \mathcal{M}_k(\mathbf{X}) \cos(t\Theta) + \Delta \sin(t\Theta), \quad \text{for } t \in [0, 1], \quad (7)$$

914 and the length of this geodesic corresponds exactly to the *projection distance* defined by

$$d_{\text{proj}}(\mathcal{M}_k(\mathbf{X}), \mathcal{M}_k(\tilde{\mathbf{X}})) = \left(\sum_{i=1}^k \theta_i^2 \right)^{1/2}. \quad (8)$$

915 Proof. Preliminaries and Definitions

- 916 1. **Grassmann Manifold** $\text{Gr}(k, n)$: The set of all k -dimensional linear subspaces of \mathbb{R}^n .
- 917 2. **Orthonormal Bases**: For a k -dimensional subspace $\mathcal{S} \subset \mathbb{R}^n$, an orthonormal basis is
 918 represented by an $n \times k$ matrix Q with columns satisfying $Q^\top Q = I_k$, where I_k is the $k \times k$
 919 identity matrix.

920 **3. Principal Angles and Vectors:** Given two subspaces \mathcal{S}_1 and \mathcal{S}_2 with orthonormal bases
 921 Q_1 and Q_2 , the principal angles $0 \leq \theta_1 \leq \theta_2 \leq \dots \leq \theta_k \leq \frac{\pi}{2}$ between them are defined
 922 recursively by

$$\cos(\theta_i) = \max_{\substack{\mathbf{u} \in \mathcal{S}_1 \\ \|\mathbf{u}\|=1}} \max_{\substack{\mathbf{v} \in \mathcal{S}_2 \\ \|\mathbf{v}\|=1}} \mathbf{u}^\top \mathbf{v}, \quad \text{subject to } \mathbf{u}^\top \mathbf{u}_j = 0, \mathbf{v}^\top \mathbf{v}_j = 0, j = 1, \dots, i-1. \quad (9)$$

923 **4. Projection Distance:** The projection distance between \mathcal{S}_1 and \mathcal{S}_2 is defined as

$$d_{\text{proj}}(\mathcal{S}_1, \mathcal{S}_2) = \left(\sum_{i=1}^k \theta_i^2 \right)^{1/2}. \quad (10)$$

924 1. Computation of the Principal Angles

925 Let $Q_1 = \mathcal{M}_k(\mathbf{X})$ and $Q_2 = \mathcal{M}_k(\tilde{\mathbf{X}})$. Both Q_1 and Q_2 are $n \times k$ matrices with orthonormal
 926 columns.

927 We construct the matrix C as follows:

$$C = Q_1^\top Q_2 \in \mathbb{R}^{k \times k}. \quad (11)$$

928 Since $Q_1^\top Q_1 = I_k$ and $Q_2^\top Q_2 = I_k$, C captures the pairwise inner products between the basis
 929 vectors of Q_1 and Q_2 .

930 We then perform the Singular Value Decomposition (SVD) of C :

$$C = U \Sigma V^\top, \quad (12)$$

931 where

- 932 • $U, V \in \mathbb{R}^{k \times k}$ are orthogonal matrices, i.e., $U^\top U = V^\top V = I_k$.
- 933 • $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$ with $\sigma_i \geq 0$.

934 The singular values σ_i of C are the cosines of the principal angles between the subspaces:

$$\sigma_i = \cos(\theta_i), \quad \theta_i \in [0, \pi/2], \quad i = 1, \dots, k. \quad (13)$$

935 This result stems from the fact that the SVD aligns the basis vectors of U and V to maximize the
 936 projections in the directions of the principal angles, which correspond to the largest cosines.

937 Since principal angles θ_i are defined in the range $[0, \pi/2]$, their cosines naturally lie in $[0, 1]$, matching
 938 the range of the singular values of C . Thus, the singular values encode the geometric relationship
 939 between the subspaces U and V in terms of the principal angles. This connection is fundamental to
 940 Grassmannian geometry, as it allows the distances and alignments between subspaces to be analyzed
 941 using the principal angles and their cosines.

942 2. Construction of Orthonormal Bases Aligned with Principal Directions

943 Define new orthonormal bases:

$$A = Q_1 U, \quad B = Q_2 V.$$

944 Verification of Orthonormality:

$$\begin{aligned} A^\top A &= (Q_1 U)^\top (Q_1 U) = U^\top Q_1^\top Q_1 U = U^\top I_k U = U^\top U = I_k, \\ B^\top B &= (Q_2 V)^\top (Q_2 V) = V^\top Q_2^\top Q_2 V = V^\top I_k V = V^\top V = I_k. \end{aligned}$$

945 We then compute $A^\top B$:

$$\begin{aligned} A^\top B &= (Q_1 U)^\top (Q_2 V) = U^\top Q_1^\top Q_2 V = U^\top C V = U^\top (U \Sigma V^\top) V \\ &= U^\top U \Sigma V^\top V = I_k \Sigma I_k = \Sigma. \end{aligned}$$

946 Thus, $A^\top B = \Sigma = \text{diag}(\cos(\theta_1), \dots, \cos(\theta_k))$.

947 **3. Decomposition of B in Terms of A and Δ**

948 We aim to express B as a linear combination of A and another orthonormal matrix Δ that is orthogonal
949 to A .

950 Let us define Δ :

$$\Delta = (B - A \cos(\Theta)) \sin(\Theta)^{-1}, \quad (14)$$

951 where $\cos(\Theta) = \Sigma$ and $\sin(\Theta) = \text{diag}(\sin(\theta_1), \dots, \sin(\theta_k))$, and $\sin(\Theta)^{-1}$ denotes the diagonal
952 matrix with entries $\sin(\theta_i)^{-1}$.

953 **Verification that Δ is Orthogonal to A :**

$$\begin{aligned} A^\top \Delta &= A^\top (B - A \cos(\Theta)) \sin(\Theta)^{-1} \\ &= (A^\top B - A^\top A \cos(\Theta)) \sin(\Theta)^{-1} \\ &= (\Sigma - I_k \cos(\Theta)) \sin(\Theta)^{-1} \\ &= (\cos(\Theta) - \cos(\Theta)) \sin(\Theta)^{-1} = 0. \end{aligned}$$

954 **Verification that Δ is Orthonormal:**

955 First, we compute $\Delta^\top \Delta$:

$$\begin{aligned} \Delta^\top \Delta &= ((B - A \cos(\Theta)) \sin(\Theta)^{-1})^\top ((B - A \cos(\Theta)) \sin(\Theta)^{-1}) \\ &= \sin(\Theta)^{-1} (B - A \cos(\Theta))^\top (B - A \cos(\Theta)) \sin(\Theta)^{-1}. \end{aligned}$$

956 We compute the inner term:

$$\begin{aligned} (B - A \cos(\Theta))^\top (B - A \cos(\Theta)) &= (B^\top - \cos(\Theta) A^\top)(B - A \cos(\Theta)) \\ &= B^\top B - B^\top A \cos(\Theta) - \cos(\Theta) A^\top B \\ &\quad + \cos(\Theta) A^\top A \cos(\Theta). \end{aligned}$$

957 Since $A^\top A = I_k$, $B^\top B = I_k$, and $A^\top B = \Sigma = \cos(\Theta)$:

$$\begin{aligned} (B - A \cos(\Theta))^\top (B - A \cos(\Theta)) &= I_k - \cos(\Theta)^\top \cos(\Theta) - \cos(\Theta)^\top \cos(\Theta) \\ &\quad + \cos(\Theta)^\top \cos(\Theta) \cos(\Theta)^\top \cos(\Theta) \\ &= I_k - \cos^2(\Theta) - \cos^2(\Theta) + \cos^4(\Theta) \\ &= I_k - 2 \cos^2(\Theta) + \cos^4(\Theta). \end{aligned}$$

958 But since $\sin^2(\Theta) = I_k - \cos^2(\Theta)$, we can write:

$$I_k - 2 \cos^2(\Theta) + \cos^4(\Theta) = (I_k - \cos^2(\Theta))^2 = \sin^4(\Theta). \quad (15)$$

959 Thus,

$$\Delta^\top \Delta = \sin(\Theta)^{-1} \sin^4(\Theta) \sin(\Theta)^{-1} = \sin^2(\Theta) I_k = I_k.$$

960 Therefore, Δ is orthonormal.

961 **Expressing B in Terms of A and Δ :**

962 Using Equation (14), we have:

$$B = A \cos(\Theta) + \Delta \sin(\Theta). \quad (16)$$

963 **4. Define the Geodesic Path**

964 On the Grassmann manifold, the geodesic $\gamma(t)$ from A to B is given by:

$$\gamma(t) = A \cos(t\Theta) + \Delta \sin(t\Theta), \quad t \in [0, 1]. \quad (17)$$

965 **Verification of Endpoints:**

966 At $t = 0$:

$$\gamma(0) = A \cos(0 \cdot \Theta) + \Delta \sin(0 \cdot \Theta) = AI_k + \Delta \cdot 0 = A.$$

967 At $t = 1$:

$$\gamma(1) = A \cos(\Theta) + \Delta \sin(\Theta) = B. \quad (18)$$

968 Thus, $\gamma(t)$ is a continuous path on $\text{Gr}(k, n)$ connecting A and B .

969 **Relate Back to Original Bases:**

970 Recall that $A = Q_1 U = \mathcal{M}_k(\mathbf{X})U$ and $B = Q_2 V = \mathcal{M}_k(\tilde{\mathbf{X}})V$.

971 Since U and V are orthogonal matrices, the subspaces spanned by Q_1 and A , and by Q_2 and B , are identical. Therefore, we can express the geodesic in terms of $\mathcal{M}_k(\mathbf{X})$ and Δ .

973 **Expressing the Geodesic in Original Terms:**

974 Let us redefine Δ accordingly to absorb U and V , so that we can write:

$$\gamma(t) = \mathcal{M}_k(\mathbf{X}) \cos(t\Theta) + \Delta \sin(t\Theta).$$

975 **5. Compute the Length of the Geodesic**

976 The length L of the geodesic $\gamma(t)$ is given by:

$$L = \int_0^1 \|\dot{\gamma}(t)\|_F dt, \quad (19)$$

977 where $\|\cdot\|_F$ denotes the Frobenius norm.

978 **Compute the Derivative $\dot{\gamma}(t)$:**

979 Since $\gamma(t) = \mathcal{M}_k(\mathbf{X}) \cos(t\Theta) + \Delta \sin(t\Theta)$, we have:

$$\dot{\gamma}(t) = -\mathcal{M}_k(\mathbf{X})\Theta \sin(t\Theta) + \Delta\Theta \cos(t\Theta),$$

980 where we used the fact that the derivative of $\cos(t\Theta)$ with respect to t is $-\Theta \sin(t\Theta)$, and similarly
981 for $\sin(t\Theta)$.

982 **Compute the Squared Norm $\|\dot{\gamma}(t)\|_F^2$:**

$$\begin{aligned} \|\dot{\gamma}(t)\|_F^2 &= \text{Tr}(\dot{\gamma}(t)^\top \dot{\gamma}(t)) \\ &= \text{Tr}\left((- \mathcal{M}_k(\mathbf{X})\Theta \sin(t\Theta) + \Delta\Theta \cos(t\Theta))^\top (- \mathcal{M}_k(\mathbf{X})\Theta \sin(t\Theta) + \Delta\Theta \cos(t\Theta))\right) \\ &= \text{Tr}\left(\Theta^2 (\sin^2(t\Theta)\mathcal{M}_k(\mathbf{X})^\top \mathcal{M}_k(\mathbf{X}) + \cos^2(t\Theta)\Delta^\top \Delta - \sin(t\Theta)\cos(t\Theta)(\mathcal{M}_k(\mathbf{X})^\top \Delta - \Delta^\top \mathcal{M}_k(\mathbf{X})))\right). \end{aligned}$$

983 Since $\mathcal{M}_k(\mathbf{X})^\top \mathcal{M}_k(\mathbf{X}) = I_k$, $\Delta^\top \Delta = I_k$, and $\mathcal{M}_k(\mathbf{X})^\top \Delta = 0$, the cross terms vanish, and we
984 have:

$$\begin{aligned} \|\dot{\gamma}(t)\|_F^2 &= \text{Tr}(\Theta^2 (\sin^2(t\Theta)I_k + \cos^2(t\Theta)I_k)) \\ &= \text{Tr}(\Theta^2 I_k) \\ &= \sum_{i=1}^k \theta_i^2. \end{aligned}$$

985 **Compute the Length L :**

986 Since $\|\dot{\gamma}(t)\|_F$ is constant with respect to t , we have:

$$\begin{aligned} L &= \int_0^1 \|\dot{\gamma}(t)\|_F dt = \|\dot{\gamma}(t)\|_F \int_0^1 dt \\ &= \left(\sum_{i=1}^k \theta_i^2 \right)^{1/2} \cdot 1 \\ &= \left(\sum_{i=1}^k \theta_i^2 \right)^{1/2}. \end{aligned}$$

987 6. Length Equals the Projection Distance

988 Comparing the computed length L with the projection distance defined in Equation (10), we find:

$$L = d_{\text{proj}}(\mathcal{M}_k(\mathbf{X}), \mathcal{M}_k(\tilde{\mathbf{X}})) = \left(\sum_{i=1}^k \theta_i^2 \right)^{1/2}. \quad (20)$$

989 On the Grassmann manifold, the geodesic distance between two subspaces is given by the length of
990 the shortest path connecting them. This distance is intrinsically linked to the principal angles between
991 the subspaces. The projection distance quantifies the separation between subspaces in terms of these
992 principal angles.

993 By computing the squared norm of the derivative of the geodesic, we find that it equals the sum of
994 the squares of the principal angles, which is the squared projection distance. Since the derivative's
995 norm is constant, the total length of the geodesic over the interval $t \in [0, 1]$ is precisely the projection
996 distance.

997 Therefore, the length of the geodesic $\gamma(t)$ connecting $\mathcal{M}_k(\mathbf{X})$ and $\mathcal{M}_k(\tilde{\mathbf{X}})$ on the Grassmann
998 manifold equals the projection distance between these two subspaces.

999 This completes the proof of Theorem 3.6.

1000

□

1001 C Proof of Theorem 3.5 - Metric Robustness

1002 **Theorem 3.5** Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_\ell] \in \mathbb{R}^{\ell \times n}$ and $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_\ell] \in \mathbb{R}^{\ell \times n}$ be two sequences of
1003 state snapshots. Suppose that both \mathbf{X} and $\tilde{\mathbf{X}}$ with $\mathcal{M}_k(\mathbf{X}) \in \mathbb{R}^{n \times k}$ and $\mathcal{M}_k(\tilde{\mathbf{X}}) \in \mathbb{R}^{n \times k}$ as the
1004 respective DMD eigenvectors, Λ and $\tilde{\Lambda}$ as the respective DMD eigenvalues, and admit a DMD form
1005 with the same initial condition $\mathbf{x}_0 = \tilde{\mathbf{x}}_0$, i.e.,

$$\forall t, \tilde{\mathbf{x}}_t = \mathcal{M}_k(\mathbf{X}) \Lambda^t \mathcal{M}_k(\mathbf{X})^\dagger \mathbf{x}_0, \quad \tilde{\mathbf{x}}_t = \mathcal{M}_k(\tilde{\mathbf{X}}) \tilde{\Lambda}^t \mathcal{M}_k(\tilde{\mathbf{X}})^\dagger \mathbf{x}_0.$$

1006 Let E_t be the difference in dynamics between \mathbf{X} and $\tilde{\mathbf{X}}$, i.e., $\forall t, \mathbf{x}_t - \tilde{\mathbf{x}}_t = E_t \mathbf{x}_0$. We have,
1007 $\forall t, \tilde{d}_{\text{proj}}(\mathcal{M}_k(\mathbf{X}), \mathcal{M}_k(\tilde{\mathbf{X}})) \leq \frac{\|E_t\|_F}{\delta_t}$ where δ_t is the spectral gap of Λ^t , and $\|\cdot\|_F$ is the Frobenius
1008 Norm.

1009 *Proof.* Let $X = [x_1, \dots, x_\ell]$, $\tilde{X} = [\tilde{x}_1, \dots, \tilde{x}_\ell] \in \mathbb{R}^{n \times \ell}$, and denote their k -dominant DMD
1010 bases by $Q := M_k(X) \in \mathbb{R}^{n \times k}$, $\tilde{Q} := M_k(\tilde{X}) \in \mathbb{R}^{n \times k}$. We define the time- t linear propagators
1011 that generate the snapshots

$$A_t := Q \Lambda^t Q^\dagger, \quad \tilde{A}_t := \tilde{Q} \tilde{\Lambda}^t \tilde{Q}^\dagger, \quad \text{so that} \quad x_t = A_t x_0, \quad \tilde{x}_t = \tilde{A}_t x_0.$$

1012 Because x_0 is arbitrary, we identify the perturbation matrix $A_t - \tilde{A}_t = E_t$.

1013 Wedin's theorem for diagonalizable matrices states that if E perturbs a matrix A whose spectrum
1014 splits into two clusters separated by a gap δ , then $\|\sin \Theta\|_F \leq \frac{\|E\|_F}{\delta}$, where Θ collects the principal
1015 angles between the invariant subspaces associated with the chosen spectral clusters.

Applying Wedin with $A = A_t$, $E = E_t$, and the dominant invariant subspace $\text{span}(Q)$, the gap is $\delta = \delta_t$. The left-hand side is exactly the Grassmann projection distance $d_{\text{proj}}(Q, \tilde{Q}) = \|\sin \Theta\|_F$. Hence $d_{\text{proj}}(Q, \tilde{Q}) \leq \frac{\|E_t\|_F}{\delta_t}$ which is the desired bound. \square

D Datasets and Implementation Details

D.1 Basic Statistics on the Datasets

Sine waves. We generated a synthetic dataset consisting of two sets of sine waves to represent a bimodal distributed data. The data were generated using the following formula:

$$y(t) = A \cdot \sin(2\pi ft + \phi), \quad (21)$$

where A is the amplitude, f is the frequency, t is the time variable and ϕ is the phase angle of the sine wave. Each mode consists of 2000 samples with phases being randomly chosen between 0 and 2π . For all the samples, the duration is 2 seconds and the sample rate is 12, making the length of each sequence be 24. $A = 0.5$ and $f = 1$ Hz for the first mode, and $A = 5$ and $f = 0.5$ Hz for the second mode.

Stock price. To test our framework on a complex multimodal dataset, we used Google stocks data from 2004 to 2019, which was used in [60]. The data consists of 6 features which are daily open, high, low, close, adjusted close, and volume. The time series were then cut into sequences with length 24, following the setup in the work done by [60].

Energy. We conducted experiments on UCI’s air quality dataset [55] consisting of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device in an Italian city. Data was recorded from March 2004 to February 2005 and consists of 28 features. Unlike the previous datasets, this one has an unimodal distribution. The data is cut into several sequences of length 7.

Electricity Transformer Temperature and humidity (ETTh). The ETTh dataset focuses on temperature and humidity data from electricity transformers [62]. It includes 2 years of data at an hourly granularity, providing detailed temporal information about transformer conditions.

Table 4 provides an overview of the datasets used in our experiments, including Sine, Stock, Energy, and ETTh. These datasets vary in both the number of samples and feature dimensions, offering a diverse evaluation setting for generative models.

Table 3: Statistics of the four datasets used in our experiments.

Dataset	Sine	Stock	Energy	ETTh
#Samples	10,000	3,773	19,711	17,420
Dimension	5	6	28	8

D.2 Baseline Metrics

We compared our proposed metric DMD-GEN with well-established time series evaluation metrics. Specifically, this comparison includes three key metrics:

Predictive Score. [60] The predictive score evaluates how well a generative model captures the temporal dynamics of the original data. It involves training a model on the generated data and assessing its performance on a real dataset. A lower predictive score indicates that the generated data contains patterns that are more representative of the temporal patterns found in the original data.

Discriminative Score. [60] The discriminative score measures the similarity between real and generated time series data by training a binary classifier to distinguish between them.

Contextual Frechet Inception Distance (context-FID). [22] Context-FID is an adaptation of the Frechet Inception Distance (FID), a metric used to assess the quality of images created by a generative model [19]. For time series, context-FID measures the similarity between the real and generated data

distributions by computing the Frechet distance between feature representations extracted from a time series feature encoder.

D.3 Implementation Details

The experiments were conducted on an NVIDIA A100 GPU. We utilized the pyDMD package¹ in Python to compute the DMD eigenvalues and eigenvectors. For generating synthetic time series, we used the original settings and the official implementation of DiffusionTS², TimeGAN³ and TimeVAE⁴.

E Datasets and Implementation Details

E.1 Basic Statistics on the Datasets

Sine waves. We generated a synthetic dataset consisting of two sets of sine waves to represent a bimodal distributed data. The data were generated using the following formula:

$$y(t) = A \cdot \sin(2\pi ft + \phi), \quad (22)$$

where A is the amplitude, f is the frequency, t is the time variable and ϕ is the phase angle of the sine wave. Each mode consists of 2000 samples with phases being randomly chosen between 0 and 2π . For all the samples, the duration is 2 seconds and the sample rate is 12, making the length of each sequence be 24. $A = 0.5$ and $f = 1$ Hz for the first mode, and $A = 5$ and $f = 0.5$ Hz for the second mode.

Stock price. To test our framework on a complex multimodal dataset, we used Google stocks data from 2004 to 2019, which was used in [60]. The data consists of 6 features which are daily open, high, low, close, adjusted close, and volume. The time series were then cut into sequences with length 24, following the setup in the work done by [60].

Energy. We conducted experiments on UCI’s air quality dataset [55] consisting of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device in an Italian city. Data was recorded from March 2004 to February 2005 and consists of 28 features. Unlike the previous datasets, this one has an unimodal distribution. The data is cut into several sequences of length 7.

Electricity Transformer Temperature and humidity (ETTh). The ETTh dataset focuses on temperature and humidity data from electricity transformers [62]. It includes 2 years of data at an hourly granularity, providing detailed temporal information about transformer conditions.

Table 4 provides an overview of the datasets used in our experiments, including Sine, Stock, Energy, and ETTh. These datasets vary in both the number of samples and feature dimensions, offering a diverse evaluation setting for generative models.

Table 4: Statistics of the four datasets used in our experiments.

Dataset	Sine	Stock	Energy	ETTh
#Samples	10,000	3,773	19,711	17,420
Dimension	5	6	28	8

E.2 Baseline Metrics

We compared our proposed metric DMD-GEN with well-established time series evaluation metrics. Specifically, this comparison includes three key metrics:

¹<https://pydmd.github.io/PyDMD/>

²<https://github.com/Y-debug-sys/Diffusion-TS>

³<https://github.com/Y-debug-sys/Diffusion-TS>

⁴<https://github.com/zzw-zwzhang/TimeGAN-pytorch>

Predictive Score. [60] The predictive score evaluates how well a generative model captures the temporal dynamics of the original data. It involves training a model on the generated data and assessing its performance on a real dataset. A lower predictive score indicates that the generated data contains patterns that are more representative of the temporal patterns found in the original data.

Discriminative Score. [60] The discriminative score measures the similarity between real and generated time series data by training a binary classifier to distinguish between them.

Contextual Frechet Inception Distance (context-FID). [22] Context-FID is an adaptation of the Frechet Inception Distance (FID), a metric used to assess the quality of images created by a generative model [19]. For time series, context-FID measures the similarity between the real and generated data distributions by computing the Frechet distance between feature representations extracted from a time series feature encoder.

E.3 Implementation Details

The experiments were conducted on an NVIDIA A100 GPU. We utilized the pyDMD package⁵ in Python to compute the DMD eigenvalues and eigenvectors. For generating synthetic time series, we used the original settings and the official implementation of DiffusionTS⁶, TimeGAN⁷ and TimeVAE⁸.

F Synthetic Generators

To evaluate the ability of DMD-GEN to detect Mode Collapse, we generate synthetic time series using two parametric functions, denoted \mathcal{G}_1 and \mathcal{G}_2 . These generators produce diverse temporal patterns by incorporating nonlinear transformations and oscillatory components. Each function is parameterized by randomly sampled variables from a uniform distribution, ensuring variability across generated samples. Below, we give the expressions of these generators,

$$\mathcal{G}_1 = \left\{ (t, x) \mapsto \frac{a}{\cosh(x + b + 3)} \times \cos((c + 2.3) \cdot t) \mid x \in [-5, 5], t \in [0, 4\pi], (a, b, c) \sim \mathcal{U} \right\},$$

$$\mathcal{G}_2 = \left\{ (t, x) \mapsto \frac{2 + a}{\cosh(x)} \times \tanh(x) \times \sin((2.8 + b) \cdot t) \mid x \in [-5, 5], t \in [0, 4\pi], (a, b, c) \sim \mathcal{U} \right\},$$

where \mathcal{U} denotes the uniform distribution over $[0, 1]$. Each time series is discretized to a length of $T = 129$ and a dimensionality of $d = 65$. Figure 3 illustrates examples of time series generated using \mathcal{G}_1 and \mathcal{G}_2 . Generator \mathcal{G}_1 produces smooth, localized wave patterns with oscillations that gradually decay in space, resulting in broader and less frequent peaks over time. In contrast, \mathcal{G}_2 generates sharper, more structured wave patterns with higher frequency oscillations

G Comparison of Generative Models Using MTopDiv Metric

As part of our ablation study, we evaluated the MTopDiv metric, originally designed for general generative models, on time series data (Table 5). The results show high variability in standard deviations, limiting meaningful comparisons and suggesting that MTopDiv is not well-suited for evaluating time series generative models.

Dataset	TimeVAE	TimeGAN	DiffusionTS
Energy	424.34 \pm 19.75	467.35 \pm 49.10	402.42 \pm 34.53 \pm
ETTh	116.23 \pm 4.96 \pm	130.85 \pm 8.80	116.51 \pm 5.52
Sines	7.72 \pm 0.18	4.92 \pm 0.32	4.53 \pm 0.13

Table 5: Comparison of different models on various datasets.

⁵<https://pydmd.github.io/PyDMD/>

⁶<https://github.com/Y-debug-sys/Diffusion-TS>

⁷<https://github.com/Y-debug-sys/Diffusion-TS>

⁸<https://github.com/zzw-zwzhang/TimeGAN-pytorch>

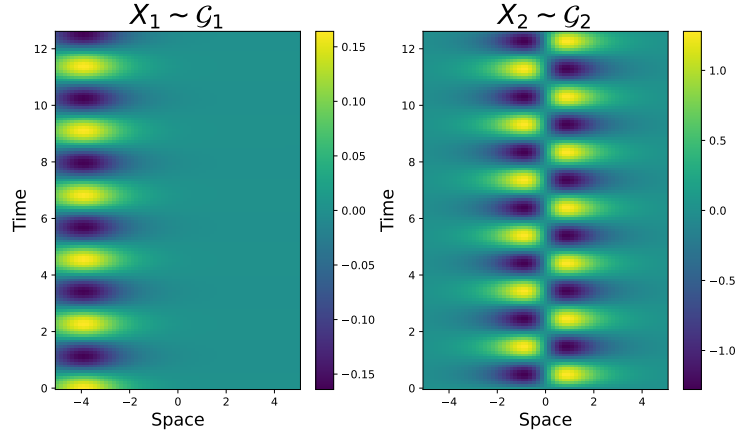


Figure 3: Examples of time series generated using the generators \mathcal{G}_1 and \mathcal{G}_2 .

1121 H Evolution of the DMD eigenvalues During Training

1122 In Figures 4, and 5, we plot the imaginary and real parts of the DMD eigenvalues of a 500 sample
 1123 original and generated time series for datasets ETTh and Sines.

1124 I Boarder impact

1125 This paper presents work whose goal is to advance the field of Machine Learning. There are many
 1126 potential societal consequences of our work, none which we feel must be specifically highlighted
 1127 here.

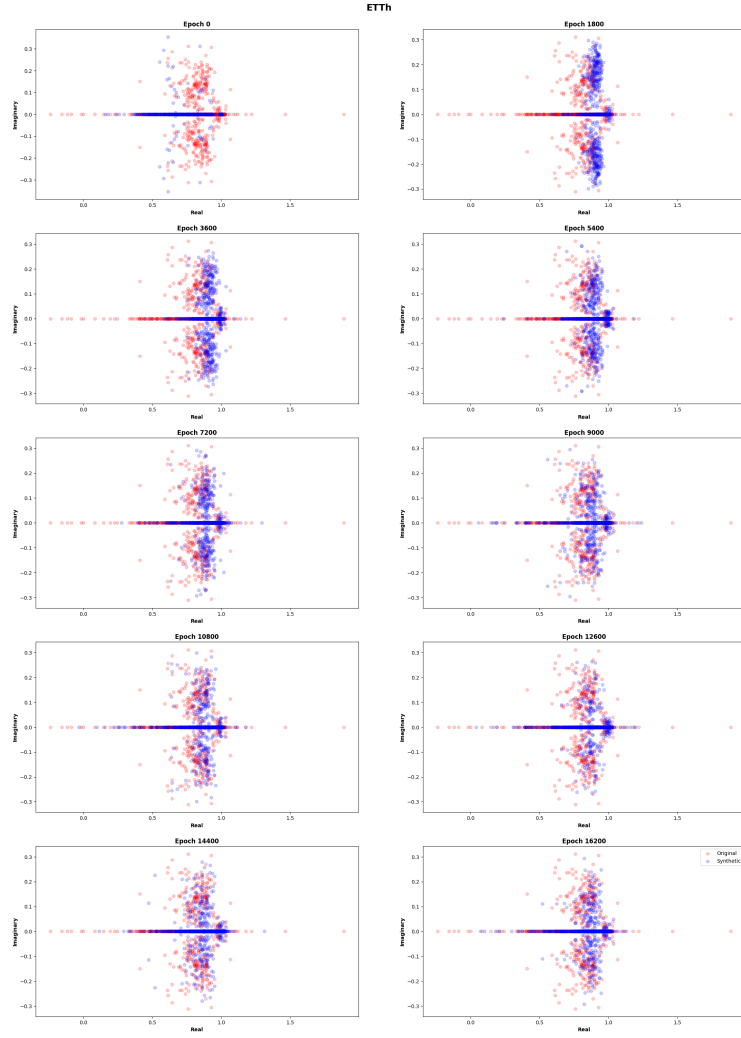


Figure 4: Comparison of DMD Eigenvalues between Original and Generated Time Series for DiffusionTS through Epochs on the dataset ETTh.

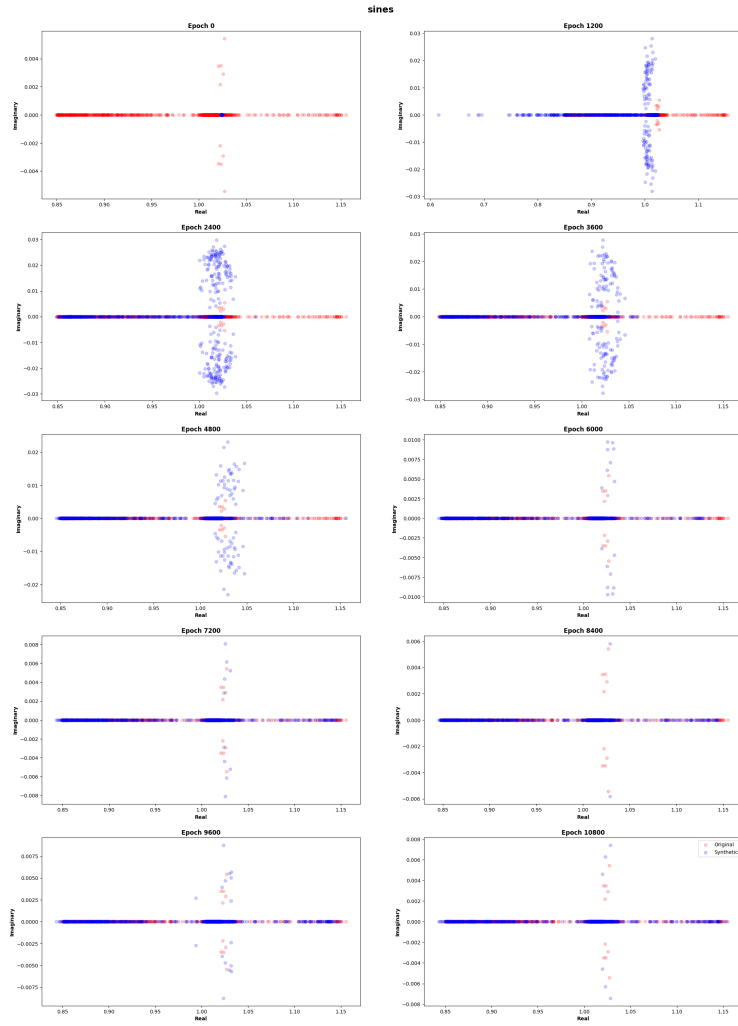


Figure 5: Comparison of DMD Eigenvalues between Original and Generated Time Series for DiffusionTS through Epochs on the dataset Sines.