BOUNDED DOMAINS FOR SYMPLECTIC GROUPS AND APPLICATIONS TO REPRESENTATION LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural networks on the Siegel disk domain have recently been introduced as an effective tool for addressing a number of machine learning tasks. These networks are built upon the complex structure of the Siegel disk domain and its connection with the real symplectic group. In this paper, we study generalizations of the real symplectic group referred to as symplectic groups over involutive algebras, and propose a class of upper half-space and precompact models associated with these groups. We derive closed-form formulas for some mappings required to implement neural networks on the precompact model, i.e., automorphisms and their inverses. We give explicit expressions for these mappings on a class of bounded domains, and investigate their representation power on the tasks of radar clutter classification, action recognition, graph reconstruction, and node classification. Our experimental results demonstrate the efficacy of the proposed approach.

1 INTRODUCTION

Riemannian neural networks have been successfully applied to address a variety of problems in domains such as natural language processing (Chami et al., 2019; Ganea et al., 2018), computer vision (Bdeir et al., 2024; van Spengler et al., 2023), and brain-computer interface (Kobler et al., 2022; Pan et al., 2022). Different families of representation spaces have been considered, e.g., hyperbolic spaces (Ganea et al., 2018), symmetric positive definite (SPD) manifolds (Huang & Gool, 2017), Grassmann manifolds (Huang et al., 2018), and manifolds of orthogonal matrices (Huang et al., 2017). These cover a wide range of spaces commonly encountered in practice.

However, in some signal processing tasks (Cabanes & Nielsen, 2021), data can be represented in spaces with rich geometric structures and do not belong to any of the spaces mentioned above. This is the case of Siegel spaces which generalize the complex Poincaré upper plane and the complex Poincaré disk (Helgason, 1979) to spaces of symmetric complex matrices (Nielsen, 2020). Recently, some approaches have been proposed to deal with learning parametric models, i.e., neural networks on Siegel spaces (López et al., 2021; Nguyen et al., 2025; Nguyen & Grozavu, 2026). The work in López et al. (2021) demonstrates the effectiveness of Siegel spaces for learning and visualizing embeddings in natural language processing and graph tasks. The recent works in Nguyen et al. (2025); Nguyen & Grozavu (2026) follow this line of research by developing discriminative Siegel neural networks. These networks are built on the Siegel upper half-space and the Siegel disk domain which are symmetric spaces associated with the real symplectic group (Helgason, 1979). There exists in the literature an extension of the real symplectic group called the Hermitian modular group (Braun, 1951). Although both these groups share many theoretical properties, the symmetric spaces associated with the Hermitian modular group seem to attract less attention than the Siegel upper half-space and the Siegel disk domain for machine learning applications.

In this paper, we study the upper half-space and bounded models associated with the Hermitian modular group. Our approach is based on the approach in Alessandrini et al. (2021); Huang et al. (2025) which concerns with symplectic groups over involutive algebras. This approach offers a general framework for studying the structure theory of several Lie groups and their associated symmetric spaces. Since our spaces of interest encompass Siegel spaces, the former have the potential to capture richer geometric structures than the latter. This may result in practical benefits, as demonstrated in López et al. (2021), where Siegel spaces achieve better performance than hyperbolic and SPD spaces on graph embedding tasks. In summary, our contributions are the following:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

- We introduce a novel family of upper half-space and precompact models associated with symplectic groups over involutive algebras. We identify group actions, and derive closed-form expressions for automorphisms and their inverses on these spaces.
- For given realizations of the proposed models that are symmetric spaces associated with the Hermitian modular group, we provide explicit expressions for the automorphisms and their inverses on the considered spaces, and propose some core components for neural networks on these spaces, e.g., batch normalization (BN) and classification layers. We also propose a method to improve graph convolutional networks (GCN) on SPD manifolds.
- We demonstrate the effectiveness of the proposed approach on the tasks of radar clutter classification, action recognition, graph reconstruction, and node classification.

2 Mathematical Background

This section recaps important concepts of the framework proposed in Alessandrini et al. (2021). See Alessandrini et al. (2021); Huang et al. (2025) for further details.

2.1 Involutive Algebras

Let $\mathbb{K}$ be a field and let A be a unital associative possibly noncommutative finite-dimensional $\mathbb{K}$-algebra. If $\mathbb{K}$ is a topological field, A has a well defined topology.

Definition 2.1. *An anti-involution on A is a $\mathbb{K}$-linear map $\sigma : A \to A$ such that $\sigma(ab) = \sigma(b)\sigma(a)$ and $\sigma^2 = \mathrm{Id}$. An involutive $\mathbb{K}$-algebra is a pair (A, σ), where A is a $\mathbb{K}$-algebra and σ is an anti-involution on A. The algebra $\mathrm{Mat}_n(A)$ of $n \times n$ matrices over A can be endowed with the anti-involution σ^T, which applies σ to each entry and then the transpose. Define*

$$A^\sigma = \{a \in A : \sigma(a) = a\}.$$

Let $A^\times$ be the group of all invertible elements of A. The set of σ-positive elements is defined as

$$A^\sigma_+ = \left\{ \sum_{l=1}^{h} a_l^2 : a_l \in (A^\sigma)^\times, h \in \mathbb{N} \right\}.$$

The set of σ-non-negative elements is defined as

$$A^\sigma_{\geq 0} = \left\{ \sum_{l=1}^{h} a_l^2 : a_l \in A^\sigma, h \in \mathbb{N} \right\}.$$

Alessandrini et al. (2021) focused on the family of involutive algebras called Hermitian algebras.

Definition 2.2. *Let $\mathbb{K}$ be a real closed field. A unital associative $\mathbb{K}$-algebra with an anti-involution (A, σ) is called Hermitian if for all $x, y \in A^\sigma$, $x^2 + y^2 = 0$ implies $x = y = 0$.*

2.2 Symplectic Groups over Involutive Algebras

Before we define symplectic groups over unital rings, let us define automorphism groups associated with σ-sesquilinear forms.

Definition 2.3. *Let A be a unital associative finite dimensional algebra over a field $\mathbb{K}$ with an anti-involution σ. We view A^2 as the set of columns and endow it with the structure of a right A-module. A σ-sesquilinear form ω on A^2 is a map*

$$\omega : A^2 \times A^2 \to A$$

such that

$$\begin{aligned} \omega(x + y, z) &= \omega(x, z) + \omega(y, z) \\ \omega(x, y + z) &= \omega(x, y) + \omega(x, z) \\ \omega(x_1 r_1, x_2 r_2) &= \sigma(r_1)\omega(x_1, x_2) r_2. \end{aligned}$$

Define

$$\mathrm{Aut}(\omega) = \{f \in \mathrm{Aut}(A^2) : \forall x, y \in A^2 : \omega(f(x), f(y)) = \omega(x, y)\}.$$

We are now ready to define symplectic groups over unital rings, a key concept used in our work.

Definition 2.4. *Let (A, σ) be a unital ring with anti-involution. Let $\Omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, and let $\omega(x, y) = \sigma(x)^T \Omega y$ be a σ-sesquilinear form on A^2. The form ω is called the standard symplectic form on A^2. The symplectic group $\mathrm{Sp}_2(A, \sigma)$ over (A, σ) is defined to be the automorphism group $\mathrm{Aut}(\omega)$. By Definition 2.3,*

$$\mathrm{Sp}_2(A, \sigma) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \sigma(a)c, \sigma(b)d \in A^\sigma, \sigma(a)d - \sigma(c)b = 1 \right\}.$$

The constructions of symplectic groups over real algebras and their associated symmetric spaces can also be extended to more general cases. To this end, we need to define complex extensions of real algebras and anti-involutions.

Definition 2.5. *Let $\mathbb{K}$ be a real closed field. We denote the algebraic closure of $\mathbb{K}$ by $\mathbb{K}_\mathbb{C} = \mathbb{K}[\mathbf{i}]$, where $\mathbf{i}$ is a square root of -1 called the imaginary unit of $\mathbb{K}_\mathbb{C}$. The field $\mathbb{K}_\mathbb{C}$ is called the complexification of $\mathbb{K}$. If $\mathbb{K} = \mathbb{R}$, then we denote $\mathbb{R}_\mathbb{C}$ as usual by $\mathbb{C}$.*

Definition 2.6. *If A is an algebra over a real closed field $\mathbb{K}$, then we call $A_\mathbb{C} = A \otimes_\mathbb{K} \mathbb{K}_\mathbb{C}$ the complexification of A.*

Definition 2.7. *Let (A, σ) be a real Hermitian algebra, and let $A_\mathbb{C} = A \otimes_\mathbb{R} \mathbb{C}$ be the complexification of A. Let $\sigma_\mathbb{C} : A_\mathbb{C} \to A_\mathbb{C}$ and $\bar{\sigma}_\mathbb{C} : A_\mathbb{C} \to A_\mathbb{C}$ be the $\mathbb{C}$-linear and $\mathbb{C}$-antilinear extensions of σ given by*

$$\sigma_\mathbb{C}(z_1 + z_2\mathbf{i}) = \sigma(z_1) + \sigma(z_2)\mathbf{i},$$
$$\bar{\sigma}_\mathbb{C}(z_1 + z_2\mathbf{i}) = \sigma(z_1) - \sigma(z_2)\mathbf{i},$$

where $z_1, z_2 \in A$.

3 Proposed Upper Half-space and Precompact Models

For the sake of clarity, in the following, we use bold uppercase letters to denote elements of the studied spaces, as these spaces consist of matrices.

We propose a new class of upper half-space and precompact models associated with $\mathrm{Sp}_2(A_\mathbb{C}, \bar{\sigma}_\mathbb{C})$. The upper half-space model is defined as

$$\mathfrak{U} = \left\{ \mathbf{Z} : \frac{1}{2\mathbf{i}}(\mathbf{Z} - \bar{\sigma}_\mathbb{C}(\mathbf{Z})) \in (A_\mathbb{C}^{\bar{\sigma}_\mathbb{C}})_+ \right\}.$$

The precompact model is defined as

$$\mathfrak{B} = \{ \mathbf{Z} : \mathbf{I} - \bar{\sigma}_\mathbb{C}(\mathbf{Z})\mathbf{Z} \in (A_\mathbb{C}^{\bar{\sigma}_\mathbb{C}})_+ \}.$$

where $\mathbf{I}$ is an identity matrix of the same size as $\mathbf{Z}$. We first present the mappings between $\mathfrak{U}$ and $\mathfrak{B}$.

Proposition 3.1. *Let $\mathbf{Z}$ and $\mathbf{Z}_0$ be two points of $\mathfrak{U}$ and $\mathfrak{B}$, respectively. Then, the following mappings convert points between $\mathfrak{U}$ and $\mathfrak{B}$:*

$$\varphi(\mathbf{Z}) = (\mathbf{Z} - \mathbf{i}\mathbf{I})(\mathbf{I} - \mathbf{i}\mathbf{Z})^{-1}, \tag{1}$$

$$\varphi^{(-1)}(\mathbf{Z}_0) = (\mathbf{Z}_0 + \mathbf{i}\mathbf{I})(\mathbf{i}\mathbf{Z}_0 + \mathbf{I})^{-1}. \tag{2}$$

The specification of mappings from each of the models into itself is crucial to construct its automorphisms (Siegel, 1943). These mappings are given in the following proposition.

Proposition 3.2. *Let $\mathbf{L} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{I} & \mathbf{i}\mathbf{I} \\ \mathbf{i}\mathbf{I} & \mathbf{I} \end{pmatrix}$, $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \in \mathrm{Sp}_2(A_\mathbb{C}, \bar{\sigma}_\mathbb{C})$, and $\mathbf{M}_0 = \mathbf{L}^{-1}\mathbf{M}\mathbf{L}$. Then, the Möbius transformation associated with $\mathbf{M}$ maps $\mathfrak{U}$ into itself, and the Möbius transformation associated with $\mathbf{M}_0$ maps $\mathfrak{B}$ into itself.*

From Proposition 3.2, one can derive mappings from $\mathfrak{B}$ to $\mathfrak{B}$ which will play a crucial role in our construction of neural networks on the proposed spaces.

Proposition 3.3. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two points of* $\mathfrak{B}$ *such that* $\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}) \in (A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$ *and* $\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y} \in A_{\mathbb{C}}^{\times}$. *Let* $\phi_{\mathbf{X}}(\mathbf{Y})$ *be defined by*

$$\phi_{\mathbf{X}}(\mathbf{Y}) = (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-\frac{1}{2}} (\mathbf{Y} - \mathbf{X})(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y})^{-1} (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}} .$$

Then

$$\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\phi_{\mathbf{X}}(\mathbf{Y}))\phi_{\mathbf{X}}(\mathbf{Y}) = (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}} (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{X})^{-1}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{Y})(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y})^{-1} (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}} .$$

Consequently, we have the following corollary.

Corollary 3.4. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two points of* $\mathfrak{B}$ *such that the conditions in Proposition 3.3 hold. Then* $\phi_{\mathbf{X}}(\mathbf{Y}) \in \mathfrak{B}$.

We can also derive the inverse map $\phi_{\mathbf{X}}^{(-1)}(\cdot)$ of $\phi_{\mathbf{X}}(\cdot)$ which is an essential transformation in BN-like layers in Siegel neural networks (Nguyen & Grozavu, 2026).

Proposition 3.5. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two points of* $\mathfrak{B}$ *such that* $\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}) \in (A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$ *and* $\mathbf{I} + \mathbf{Y}f(\mathbf{X}) \in A_{\mathbb{C}}^{\times}$, *where* $f(\mathbf{X}) = (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}} \bar{\sigma}_{\mathbb{C}}(\mathbf{X}) (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}$. *Let* $g(\mathbf{X}) = (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-\frac{1}{2}} \mathbf{X} (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}}$. *Then, the function* $\phi_{\mathbf{X}}^{(-1)}(\mathbf{Y})$ *is given by*

$$\phi_{\mathbf{X}}^{(-1)}(\mathbf{Y}) = (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}} (\mathbf{I} + \mathbf{Y}f(\mathbf{X}))^{-1}(\mathbf{Y} + g(\mathbf{X})) (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}} .$$

The proposed upper half-space model has a close connection with the set of $\bar{\sigma}_{\mathbb{C}}$-positive elements. This connection is established in the following proposition.

Proposition 3.6. *Let* $\mathbf{Z} \in A_{\mathbb{C}}$, $\mathbf{X}_0 = \frac{1}{2}(\mathbf{Z} + \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}))$, $\mathbf{Y}_0 = \frac{1}{2\mathbf{i}}(\mathbf{Z} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}))$ *such that* $\mathbf{Y}_0 \in A_{\mathbb{C}}^{\times}$, *and let* ρ *be the mapping defined by*

$$\rho(\mathbf{Z}) = \begin{pmatrix} \mathbf{Y}_0 + \mathbf{X}_0\mathbf{Y}_0^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}_0) & \mathbf{X}_0\mathbf{Y}_0^{-1} \\ \mathbf{Y}_0^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}_0) & \mathbf{Y}_0^{-1} \end{pmatrix}.$$

Then $\mathbf{Z} \in \mathfrak{U}$ *if and only if* $\rho(\mathbf{Z}) \in (A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$.

4 Realizations of the Proposed Models

We now consider realizations of the models introduced in Section 3 where $A_{\mathbb{C}} = \mathrm{Mat}_n(\mathbb{R}) \oplus \mathbb{C}$. In this case, the group $\mathrm{Sp}_2(A_{\mathbb{C}}, \bar{\sigma}_{\mathbb{C}})$ is isomorphic to the Hermitian modular group (Braun, 1951). The upper half-space $\mathbb{HU}_n$ is defined as

$$\mathbb{HU}_n = \left\{ \mathbf{Z} : \frac{1}{2\mathbf{i}}(\mathbf{Z} - \mathbf{Z}^H) \in \mathbb{H}_n^+ \right\},$$

where $\mathbf{Z}^H$ denotes the conjugate transpose of $\mathbf{Z}$, and $\mathbb{H}_n^+$ denotes the manifold of $n \times n$ Hermitian positive definite (HPD) matrices. The precompact model (hereafter referred to as a bounded domain) $\mathbb{HB}_n$ is defined as

$$\mathbb{HB}_n = \{\mathbf{Z} : \mathbf{I} - \mathbf{Z}^H\mathbf{Z} \in \mathbb{H}_n^+\}.$$

It has been shown (Braun, 1951) that the Möbius transformation associated with each point of $\mathrm{Sp}_2(A_{\mathbb{C}}, \bar{\sigma}_{\mathbb{C}})$ is an automorphism of $\mathbb{HB}_n$. The following corollaries are consequences of Propositions 3.3 and 3.5.

Corollary 4.1. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two points on* $\mathbb{HB}_n$. *Then, an automorphism* $\phi_{\mathbf{X}}(\mathbf{Y})$ *of* $\mathbb{HB}_n$ *can be given by*

$$\phi_{\mathbf{X}}(\mathbf{Y}) = (\mathbf{I} - \mathbf{X}\mathbf{X}^H)^{-\frac{1}{2}}(\mathbf{Y} - \mathbf{X})(\mathbf{I} - \mathbf{X}^H\mathbf{Y})^{-1}(\mathbf{I} - \mathbf{X}^H\mathbf{X})^{\frac{1}{2}}. \tag{3}$$

For the considered domain, the inverse map $\phi_{\mathbf{X}}^{(-1)}(\cdot)$ of the automorphism $\phi_{\mathbf{X}}(\cdot)$ has a simple form.

Corollary 4.2. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two points on* $\mathbb{HB}_n$. *Then, the inverse map* $\phi_{\mathbf{X}}^{(-1)}(\mathbf{Y})$ *of the automorphism* $\phi_{\mathbf{X}}(\mathbf{Y})$ *is given by*

$$\phi_{\mathbf{X}}^{(-1)}(\mathbf{Y}) = (\mathbf{I} - \mathbf{X}\mathbf{X}^H)^{\frac{1}{2}}(\mathbf{I} + \mathbf{Y}\mathbf{X}^H)^{-1}(\mathbf{Y} + \mathbf{X})(\mathbf{I} - \mathbf{X}^H\mathbf{X})^{-\frac{1}{2}}. \tag{4}$$

Algorithm 1 Computation of the Proposed Distance Function on $\mathbb{HIB}_n$

Input: Two points $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathbb{HIB}_n$.
$\mathbf{Z}_3 \leftarrow \varphi^{(-1)}(\mathbf{Z}_1)$
$\mathbf{Z}_4 \leftarrow \varphi^{(-1)}(\mathbf{Z}_2)$
$\mathbf{X}_0 \leftarrow \frac{1}{2}(\mathbf{Z}_3 + \mathbf{Z}_3^H)$
$\mathbf{Y}_0 \leftarrow \frac{1}{2i}(\mathbf{Z}_3 - \mathbf{Z}_3^H)$
$\mathbf{Z} \leftarrow \mathbf{Y}_0^{-\frac{1}{2}}(\mathbf{Z}_4 - \mathbf{X}_0)\mathbf{Y}_0^{-\frac{1}{2}}$
$\mathbf{Z}_0 \leftarrow \varphi(\mathbf{Z})$
Use SVD to decompose $\mathbf{Z}_0 = \mathbf{U}\mathbf{D}\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, and $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ is a real diagonal matrix
for $i = 1$ **to** n **do**
  $v_i = \log\left(\frac{1+d_i}{1-d_i}\right)$
end for
$s = \sum_{i=1}^{n} v_i$
Output: The distance s between $\mathbf{Z}_1$ and $\mathbf{Z}_2$.

5 Applications

In Section 5.1, we present some core components for building neural networks on $\mathbb{HIB}_n$. A method for improving GCNs on SPD manifolds is proposed in Section 5.2.

5.1 Building Blocks for Neural Networks on $\mathbb{HIB}_n$

5.1.1 Distance Function

We propose a distance function based on the concept of vector-valued distance (Kapovich et al., 2005; 2017; López et al., 2021). The computation of the distance function is presented in Algorithm 1, where $\log(\cdot)$ denotes the logarithmic function.

5.1.2 BN Layers

BN-like layers for neural networks on $\mathbb{HIB}_n$ can be constructed using an approach similar to Nguyen & Grozavu (2026). One of the key operations performed by these layers is the update of the running mean which relies on the concept of almost geodesics. For our BN layers, the almost geodesic $\gamma_{\mathbf{X},\mathbf{Y}} : [0,1] \to \mathbb{HIB}_n$ joining two points $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbb{HIB}_n$ is given as

$$\gamma_{\mathbf{X},\mathbf{Y}}(t) = \phi_{\mathbf{X}}^{(-1)}(\gamma_{\mathbf{0}_n,\phi_{\mathbf{X}}(\mathbf{Y})}(t)) = \phi_{\mathbf{X}}^{(-1)}(\alpha_{\mathbf{X},\mathbf{Y}}(t)\phi_{\mathbf{X}}(\mathbf{Y})), \tag{5}$$

where $\mathbf{0}_n$ is the $n \times n$ zero matrix, and $\alpha_{\mathbf{X},\mathbf{Y}} : [0,1] \to [0,1]$ is a solution of the following equation:

$$d^K(\mathbf{0}_n, \alpha_{\mathbf{X},\mathbf{Y}}(t)\phi_{\mathbf{X}}(\mathbf{Y})) = t d^K(\mathbf{0}_n, \phi_{\mathbf{X}}(\mathbf{Y})), \tag{6}$$

where $d^K(\cdot,\cdot)$ denotes the Kobayashi pseudodistance (Franzoni & Vesentini, 1980) on $\mathbb{HIB}_n$.

In Nguyen & Grozavu (2026), the effectiveness of almost geodesics was established only empirically. Here, we provide a characterization of such curves that serves as a theoretical justification for their use in our BN layers.

Proposition 5.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two points on $\mathbb{HIB}_n$, and let $\gamma_{\mathbf{X},\mathbf{Y}} : [0,1] \to \mathbb{HIB}_n$ be the almost geodesic joining these points constructed from Eqs. (5) and (6). Then $\gamma_{\mathbf{X},\mathbf{Y}}(t)$ is the point at parameter t along a minimizing geodesic from $\mathbf{X}$ to $\mathbf{Y}$.*

Proposition 5.1 suggests that each update of the running mean results in a new estimate that captures essential statistics of both the batch mean and the current running mean. The details of the BN layers on $\mathbb{HIB}_n$ are provided in Appendix C.1.

5.1.3 Classification Layers

Based on Proposition 3.6, a classification layer for neural networks on $\mathbb{HIB}_n$ can be built as follows.

Definition 5.2. *Let $\mathbf{Z} \in \mathbb{HB}_n$ be an input of the classification layer. Let $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathrm{Mat}_n(\mathbb{R})$ be the real and imaginary parts of $\varphi^{(-1)}(\mathbf{Z}) \in \mathbb{HU}_n$. Let $\tilde{\mathbf{Z}}_1 = \frac{1}{2}\left((\mathbf{Z}_1 + \mathbf{Z}_1^T) + (\mathbf{Z}_2 - \mathbf{Z}_2^T)\mathbf{i}\right)$ and $\tilde{\mathbf{Z}}_2 = \frac{1}{2}\left((\mathbf{Z}_2 + \mathbf{Z}_2^T) + (\mathbf{Z}_1^T - \mathbf{Z}_1)\mathbf{i}\right)$. Define*

$$\mathbf{X} = \psi(\mathbf{Z}) = \begin{pmatrix} \tilde{\mathbf{Z}}_2 + \tilde{\mathbf{Z}}_1\tilde{\mathbf{Z}}_2^{-1}\tilde{\mathbf{Z}}_1^H & \tilde{\mathbf{Z}}_1\tilde{\mathbf{Z}}_2^{-1} \\ \tilde{\mathbf{Z}}_2^{-1}\tilde{\mathbf{Z}}_1^H & \tilde{\mathbf{Z}}_2^{-1} \end{pmatrix}. \tag{7}$$

Then the probability of class l is computed as

$$p(y = l|\mathbf{Z}) \propto \exp\left(\frac{|\langle \log(\mathbf{P}^{-\frac{1}{2}}\mathbf{X}\mathbf{P}^{-\frac{1}{2}}), \mathbf{P}^{-\frac{1}{2}}\mathbf{W}\mathbf{P}^{-\frac{1}{2}}\rangle_F|}{\|\mathbf{P}^{-\frac{1}{2}}\mathbf{W}\mathbf{P}^{-\frac{1}{2}}\|_F}\right), \tag{8}$$

where $\mathbf{P} \in \mathbb{H}^+_{2n}$ and $\mathbf{W} \in T_{\mathbf{P}}\mathbb{H}^+_{2n}$ are the parameters of the layer, $\exp(\cdot)$ denotes the exponential function, $\log(\cdot)$ denotes (by abuse of notation) the matrix logarithm, $|\cdot|$ denotes the absolute value, $\langle\cdot,\cdot\rangle_F$ and $\|\cdot\|_F$ denote the Frobenius inner product and Frobenius norm, respectively.

The argument of the exponential function in Eq. (8) corresponds to the distance from a point to a hyperplane on $\mathbb{H}^+_{2n}$, which is obtained by extending the approaches in Lebanon & Lafferty (2004); Ganea et al. (2018); Nguyen (2022); Nguyen & Yang (2023) to HPD manifolds. The details of this extension are provided in Appendix C.2.

5.2 Improving GCNs on SPD Manifolds

In this section, we propose a method to improve GCNs on SPD manifolds (Zhao et al., 2023). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$, $\mathbf{z}_j^l, j \in \mathcal{V}$ be the embedding of node j at layer l ($l = 0$ indicates input node features), $\mathcal{N}(j) = \{k : (j,k) \in \mathcal{E}\}$ be the set of neighbors of $j \in \mathcal{V}$, $\mathbf{W}^l$ and $\mathbf{b}^l$ be the weight and bias at layer l, v_{jk} be the relative importance of node k to node j, and δ be a non-linear activation function. A basic GCN message-passing update can be written as

$$\begin{aligned} \mathbf{p}_j^l &= \mathbf{W}^l\mathbf{z}_j^{l-1} && \text{(feature transform)} \\ \mathbf{q}_j^l &= \sum_{k\in\mathcal{N}(j)} v_{jk}\mathbf{p}_k^l && \text{(aggregation)} \\ \mathbf{z}_j^l &= \delta(\mathbf{q}_j^l + \mathbf{b}^l) && \text{(bias and nonlinearity)} \end{aligned}$$

For our network, the original features of node j are first mapped to a symmetric matrix $\mathbf{X} \in \mathrm{Mat}_n(\mathbb{R})$ which is then converted to $\mathbb{HU}_n$ as

$$\mathbf{Z} = \mathbf{X} + \exp(\mathbf{X})\mathbf{i},$$

where $\exp(\cdot)$ denotes (by abuse of notation) the matrix exponential. The matrix $\mathbf{Z}$ is transformed as

$$\mathbf{Z}' = \psi(\phi_{\mathbf{P}}(\varphi(\mathbf{Z}))), \tag{9}$$

where $\mathbf{P} \in \mathbb{HB}_n$ is a parameter of our network, $\phi_{\mathbf{P}}(\cdot)$ and $\psi(\cdot)$ are the mappings defined in Eqs. (3) and (7), respectively. Then the initial embedding $\mathbf{Y}_j^0$ of node j is obtained by applying a BiMap layer (Huang & Gool, 2017) to the real embedding $\mathbf{Z}''$ of $\mathbf{Z}'$ given by

$$\mathbf{Z}'' = \begin{pmatrix} \mathrm{Re}(\mathbf{Z}') & -\mathrm{Im}(\mathbf{Z}') \\ \mathrm{Im}(\mathbf{Z}') & \mathrm{Re}(\mathbf{Z}') \end{pmatrix},$$

where $\mathrm{Re}(\mathbf{Z}')$ and $\mathrm{Im}(\mathbf{Z}')$ denote the real and imaginary parts of $\mathbf{Z}'$, respectively. Denote by $\mathbf{M}^l$ an isometry map (Zhao et al., 2023) of SPD manifolds at layer l. Let $\mathrm{Exp}_{\mathbf{I}}(\cdot)$ and $\mathrm{Log}_{\mathbf{I}}(\cdot)$ be the SPD exponential and logarithmic maps (Pennec, 2006) at the identity matrix $\mathbf{I}$, and let $v_{jk} = |\mathcal{N}(j)|^{-\frac{1}{2}}|\mathcal{N}(k)|^{-\frac{1}{2}}$. Let $\oplus$ be the binary operation defined by $\mathbf{P} \oplus \mathbf{Q} = \mathbf{P}^{\frac{1}{2}}\mathbf{Q}\mathbf{P}^{\frac{1}{2}}$ for any two points $\mathbf{P}$ and $\mathbf{Q}$ on an SPD manifold, and let $\delta(\cdot)$ be the non-linear activation function (Huang et al., 2018; Zhao et al., 2023) given by

$$\delta(\mathbf{P}) = \mathbf{U}\,\mathrm{diag}(\max(\mathbf{e}, \epsilon))\mathbf{U}^{-1},$$

Table 1: Radar clutter classification results (mean accuracy ± standard deviation) across five runs. The tuple (C, N) below each dataset indicates the number of classes C and the size of the dataset N. The dimension of the time series is set to 30.

Method	Dataset 1 $(20, 950)$	Dataset 2 $(40, 450)$	Dataset 3 $(60, 650)$	Dataset 4 $(80, 900)$	Dataset 5 $(100, 1000)$	Dataset 6 $(120, 1200)$
ComplexLSTM	26.00±3.55	24.16±1.55	22.45±0.71	46.87±2.24	17.84±0.96	27.95±2.24
kNN (Cabanes & Nielsen, 2021)	28.16±0.0	31.81±0.0	36.15±0.0	51.33±0.0	30.68±0.0	29.61±0.0
Kernel-Siegel (Chevallier et al., 2016)	27.83±0.0	33.05±0.0	35.01±0.0	52.15±0.0	32.80±0.0	28.31±0.0
SiegelNetFC (Nguyen et al., 2025)	31.24±0.44	41.16±0.55	40.02±0.48	62.47±0.51	42.86±0.22	34.22±0.46
SiegelNetBN (Nguyen & Grozavu, 2026)	33.83±0.34	43.29±0.40	41.38±0.37	65.93±0.32	45.61±0.12	37.10±0.26
HBNetBN (Ours)	**43.01±0.09**	**46.73±0.46**	**45.50±0.42**	**68.84±0.27**	**46.94±0.10**	**38.81±0.32**

Table 2: Action recognition results (mean accuracy ± standard deviation) across five runs. XSubject corresponds to the cross-subject setting of NTU60 and NTU120. XView and XSetup correspond to the cross-view and cross-setup settings of NTU60 and NTU120, respectively. For NTU60, we use the first 20 training/testing sequences of each action class. For NTU120, we use the first 10 training/testing sequences of each action class.

Method	SBU		NTU60		NTU120	
	Split 1	Split 2	XSubject	XView	XSubject	XSetup
SPDNetBN (Brooks et al., 2019)	72.53±1.31	64.69±0.29	44.41±1.10	54.33±1.59	27.89±0.29	32.54±0.79
LieBN-AIM-(1.5) (Chen et al., 2024)	72.40±1.27	63.36±1.41	43.41±0.92	53.33±1.46	28.32±1.43	33.13±0.53
LieBN-LCM-(0.5) (Chen et al., 2024)	69.66±4.09	58.54±2.13	41.91±0.89	52.24±1.13	25.09±1.12	32.07±1.42
SiegelNetBN (Nguyen & Grozavu, 2026)	**76.80±0.68**	61.61±0.75	44.75±1.09	55.26±0.80	30.60±0.45	33.26±0.62
HBNetBN (Ours)	75.52±0.56	**68.12±1.06**	**46.28±0.70**	**57.01±0.58**	**32.63±0.65**	**34.02±0.59**

where $\mathbf{P} = \mathbf{U}\,\mathrm{diag}(\mathbf{e})\mathbf{U}^{-1}$ is the eigenvalue decomposition of $\mathbf{P}$, and ϵ is a small positive constant. Let $\mathbf{B}^l$ be the SPD parameter representing the bias at layer l. Then the message-passing update of our network can be summarized as

$$\mathbf{P}_j^l = \mathbf{M}^l \odot \mathbf{Y}_j^{l-1} \qquad \text{(feature transform)}$$

$$\mathbf{Q}_j^l = \mathrm{Exp}_{\mathbf{I}}\Big(\sum_{k\in\mathcal{N}(j)} v_{jk}\,\mathrm{Log}_{\mathbf{I}}(\mathbf{P}_k^l)\Big) \qquad \text{(aggregation)}$$

$$\mathbf{Y}_j^l = \delta(\mathbf{B}^l \oplus \mathbf{Q}_j^l) \qquad \text{(bias and nonlinearity)}$$

6 Experiments

6.1 Radar Clutter Classification

We evaluate our method on the radar clutter classification task using the synthetic datasets in Nguyen & Grozavu (2026). The task aims at recognizing different types of radar clutter using information recorded by a radar related to seas, forests, fields, cities and other environmental elements surrounding the radar (Cabanes, 2022). The radar signals are stationary centered autoregressive (AR) Gaussian time series (Barbaresco, 1996; 2013; Billingsley, 2002). Following Cabanes (2022); Nguyen & Grozavu (2026), we represent the input data for our network in product space $\mathrm{Sym}_n^+ \times \mathbb{SD}_n^{r-1}$, where r ($r > 1$) is the order of the AR model, and $\mathbb{SD}_n$ is the Siegel disk domain of $n \times n$ matrices. Note that these input data also belong to product space $\mathrm{Sym}_n^+ \times \mathbb{HB}_n^{r-1}$. Our network termed HBNetBN consists of a BN layer and a classification layer. Given a batch of k points $\{(\mathbf{P}_j^0, \mathbf{X}_j^1, \ldots, \mathbf{X}_j^{r-1})\}_{j=1}^k$, where $(\mathbf{P}_j^0, \mathbf{X}_j^1, \ldots, \mathbf{X}_j^{r-1}) \in \mathrm{Sym}_n^+ \times \mathbb{HB}_n^{r-1}$, the BN layer transforms each batch of k points $\{\mathbf{X}_j^l\}_{j=1}^k, l = 1, \ldots, r-1$ to produce a batch of k points $\{(\mathbf{P}_j^0, \tilde{\mathbf{X}}_j^1, \ldots, \tilde{\mathbf{X}}_j^{r-1})\}_{j=1}^k$. Since the output of the BN layer belong to a product space, the classification layer of our network is an extension of the proposed classification layer to the product

Table 3: Graph reconstruction results (mean distortion $\pm$ standard deviation) across five runs. S and B denote the models built on the Siegel upper half-space and the Siegel disk domain, respectively. R and F_1 correspond to two types of vector-valued distance, i.e., Riemannian distance and Finsler distance inducing the l^1-metric (see López et al. (2021) for details). The subscripts indicate the embedding dimensions.

Space	Grid	Tree	Tree $\times$ Grid	Grid $\times$ Tree	Tree $\diamond$ Grid	Grid $\diamond$ Tree
SPD_6^R	14.33±0.04	5.45±0.06	9.07±0.02	8.97±0.02	2.23±0.04	5.31±0.03
$\mathrm{S}_4^{F_1}$	3.34±0.08	2.02±0.13	2.64±0.07	2.67±0.04	1.08±0.09	2.33±0.06
$\mathrm{B}_4^{F_1}$	2.92±0.12	1.59±0.06	2.61±0.06	2.63±0.03	**0.86±0.04**	1.47±0.08
$\mathbb{HB}_4^{F_1}$ (Ours)	**1.88±0.09**	**1.42±0.05**	**2.45±0.04**	**2.48±0.03**	0.97±0.03	**1.42±0.06**

Table 4: Node classification results (mean accuracy $\pm$ standard deviation) across ten runs.

Model	Airport	Pubmed	Cora	Disease
SPD-GCN (Zhao et al., 2023)	82.6±1.5	78.7±0.5	82.3±0.5	96.8±3.1
HB-GCN (Ours)	**91.1±0.7**	**79.2±0.4**	**82.6±0.5**	**97.2±1.4**

space setting (Nguyen & Yang, 2023) (see Appendix C.3). Results in Tab. 1 show that our network outperforms its competitors in terms of mean accuracy on all the datasets.

6.2 Action Recognition

We conduct action recognition experiments using three datasets, i.e., SBU (Yun et al., 2012), NTU60 (Shahroudy et al., 2016), and NTU120 (Liu et al., 2019). Our network HBNetBN is compared against SiegelNetBN (Nguyen & Grozavu, 2026) and some state-of-the-art SPD neural networks (Brooks et al., 2019; Chen et al., 2024). We consider a challenging setting in which the raw input sequences contain only the 3D joint positions of the right elbow, right hand, left elbow, and left hand of the first person. The input SPD matrices of the SPD neural networks are obtained as in Huang & Gool (2017). The corresponding input of each of these matrices $\mathbf{X}$ for HBNetBN is computed as $\log(\mathbf{X}) + \mathbf{Xi}$. Note that the classification layer of HBNetBN for these experiments is the one proposed in Section 5.1.3. Results in Tab. 2 show that HBNetBN is among the two best performers on all the datasets. Moreover, HBNetBN achieves higher mean accuracies than SiegelNetBN on five of the six datasets.

6.3 Graph Reconstruction

Following López et al. (2021), we generate datasets consisting of graphs with various geometric features, i.e., grids, trees, Cartesian and rooted products of grids and trees. We embed graph nodes in a transductive setting (Chami et al., 2019) using low-dimensional embeddings. We optimize the loss proposed in Gu et al. (2019) which captures the average distortion. The loss is constructed based on the distance function in Algorithm 1. Our model is compared against those built on SPD and Siegel spaces which are strong baselines considered in López et al. (2021). For each model, we report its best mean accuracies obtained with different vector-valued distances (López et al., 2021). Results in Tab. 3 show that the model built on $\mathbb{HB}_4$ achieves lower mean distortions than its counterparts built on Sym_6^+, $\mathbb{SD}_4$, and $\mathbb{SH}_4$ (the Siegel upper half-space of 4×4 matrices) in most cases. This indicates that the graph reconstruction task can benefit from the rich geometric structure of the studied domains.

6.4 Node Classification

We now turn to a graph downstream task. We conduct node classification experiments on four datasets, i.e., Airport (Zhang & Chen, 2018), Pubmed (Namata et al., 2012a), Cora (Sen et al., 2008), and Disease (Chami et al., 2019). We follow the inductive setting (Chami et al., 2019) and use original node features for fair comparison (Zhao et al., 2023). Our model termed HB-GCN (see Section 5.2) is compared against SPD-GCN (Zhao et al., 2023). HB-GCN and SPD-GCN learn the

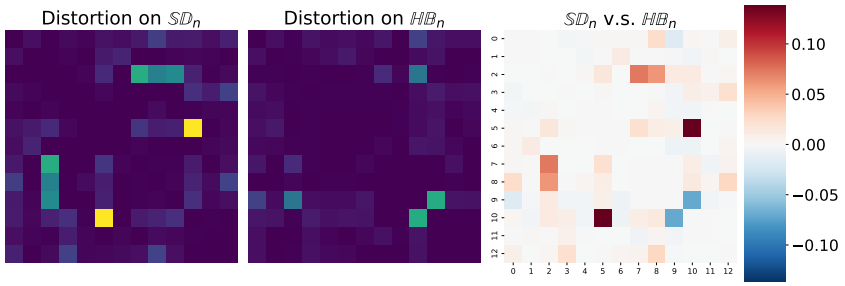

Figure 1: Comparison of reconstruction distortions on $\mathbb{SD}_n$ and $\mathbb{HB}_n$.

Table 5: Impact of automorphisms on radar clutter classification performance.

Method	Dataset 1 $(20, 950)$	Dataset 2 $(40, 450)$	Dataset 3 $(60, 650)$	Dataset 4 $(80, 900)$	Dataset 5 $(100, 1000)$	Dataset 6 $(120, 1200)$
HBClassifier	30.02±0.28	39.17±0.50	38.96±0.39	59.43±0.36	40.15±0.18	33.06±0.35
HBNet	**39.44±0.21**	**43.12±0.63**	**42.74±0.56**	**64.14±0.33**	**43.67±0.15**	**36.46±0.38**

node embeddings on $\mathbb{HB}_8$ and Sym^+_{14}, respectively. Note that in this setting, HB-GCN has fewer parameters than SPD-GCN (see Appendix A.4.3). Results in Tab. 4 show that mapping the original node features to the studied domains benefits node classification. Specifically, HB-GCN improves the mean accuracy by more than 8% with respect to SPD-GCN on Airport dataset.

6.5 Ablation Study

Representation power of $\mathbb{HB}_n$ vs. $\mathbb{SD}_n$ Since $\mathbb{SD}_n$ is a subdomain of $\mathbb{HB}_n$, we will investigate whether the additional degrees of freedom provided by $\mathbb{HB}_n$ can be exploited in our experiments to improve performance. To this end, we use the graph reconstruction task to illustrate the difference in the representation capabilities of these domains. Here we focus on comparing the distortions for all pairs of connected nodes obtained by embedding nodes in $\mathbb{HB}_n$ and $\mathbb{SD}_n$. Our experiment is conducted on a tree-like graph with 13 nodes. Fig. 1 shows the distortion matrices of $\mathbb{SD}_n$ (left panel) and $\mathbb{HB}_n$ (middle panel), as well as their difference (right panel), obtained by subtracting the distortion matrix of $\mathbb{SD}_n$ from that of $\mathbb{HB}_n$. The distortion matrices of $\mathbb{SD}_n$ and $\mathbb{HB}_n$ are averaged over five runs. As can be observed, embedding nodes in $\mathbb{HB}_n$ results in smaller distortions than those obtained by embedding nodes in $\mathbb{SD}_n$ for most pairs of connected nodes. This suggests that the richer geometric structure of $\mathbb{HB}_n$ compared to $\mathbb{SD}_n$ is exploited to improve graph reconstruction performance. This may also help explain the better performance of our networks compared to Siegel neural networks on the other tasks considered in our experiments.

Impact of automorphisms We investigate the effectiveness of automorphisms on our model's performance in the radar clutter classification experiments. To focus on the impact of automorphisms, we create a network termed HBNet by replacing the BN layer in HBNetBN with a layer based on the automorphism given in Eq. (3). We then compare HBNet to the network termed HBClassifier which consists of only the classification layer of HBNet. Results in Tab. 5 show that adding the automorphism layer consistenly improves the performance of HBClassifier across all the datasets.

7 Conclusion

We have proposed a novel class of complex spaces. We have derived closed-form expressions for automorphisms and their inverses on these spaces. We have discussed realizations of these spaces, and provide explicit expressions for the associated automorphisms and their inverses. We have developed some fundamental building blocks for neural networks on the considered spaces. Finally, we have presented experimental evaluations demonstrating the effectiveness of our approach in the tasks of radar clutter classification, action recognition, graph reconstruction, and node classification.

AI USE STATEMENT

We used AI tools for our research on related works and bibliographic entries for inclusion in our .bib file.

REFERENCES

Daniele Alessandrini, Arkady Berenstein, Vladimir Retakh, Eugen Rogozinnikov, and Anna Wienhard. Symplectic Groups over Noncommutative Algebras. *Selecta Mathematica*, 28, 2021.

R. M. Anderson and R. M. May. Infectious Diseases of Humans: Dynamics and Control. *Epidemiology and Infection*, 108(1):211–211, 1992.

Frédéric Barbaresco. Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median. *Matrix Information Geometry*, pp. 199–255, 2013.

Frdric Barbaresco. Super-resolution Spectrum Analysis Regularization: Burg, Capon and AGO-antagonistic Algorithms. In *EUSIPCO-96*, pp. 1–4, 1996.

Ahmad Bdeir, Kristian Schwethelm, and Niels Landwehr. Fully Hyperbolic Convolutional Neural Networks for Computer Vision. In *ICLR*, 2024.

J. Barrie Billingsley. *Low-angle Radar Land Clutter, Measurements and Empirical Models*. William Andrew Publishing, 2002.

Hel Braun. Hermitian Modular Functions. III: The Hermitian Modular Group. *Annals of Mathematics*, 53(1):143–160, 1951.

Daniel A. Brooks, Olivier Schwander, Frédéric Barbaresco, Jean-Yves Schneider, and Matthieu Cord. Riemannian Batch Normalization for SPD Neural Networks. In *NeurIPS*, pp. 15463–15474, 2019.

Yann Cabanes. *Multidimensional Complex Stationary Centered Gaussian Autoregressive Time Series Machine Learning in Poincaré and Siegel Disks: Application for Audio and Radar Clutter Classification*. Theses, 2022.

Yann Cabanes and Frank Nielsen. Classification in the Siegel Space for Vectorial Autoregressive Data. In *Geometric Science of Information: 5th International Conference*, pp. 693–700, 2021.

Ines Chami, Rex Ying, Christopher R, and Jure Leskovec. Hyperbolic Graph Convolutional Neural Networks. *CoRR*, abs/1910.12933, 2019. URL https://arxiv.org/abs/1910.12933.

Ziheng Chen, Yue Song, Yunmei Liu, and Nicu Sebe. A Lie Group Approach to Riemannian Batch Normalization. *CoRR*, abs/2403.11261, 2024. URL http://arxiv.org/abs/2403.11261.

Emmanuel Chevallier, Thibault Forget, Frédéric Barbaresco, and Jesus Angulo. Kernel Density Estimation on the Siegel Space with an Application to Radar Processing. *Entropy*, 18(11), 2016.

Tullio Franzoni and Edoardo Vesentini. *Holomorphic Maps and Invariant Distances*. North-Holland mathematics studies ; 40. North-Holland Pub. Co., 1980.

Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *NeurIPS*, pp. 5350–5360, 2018.

Albert Gu, Frederic Sala, Beliz Gunel, and Christopher Ré. Learning Mixed-Curvature Representations in Products of Model Spaces. In *ICLR*, 2019.

S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*. ISSN. Elsevier Science, 1979.

Pengfei Huang, Georgios Kydonakis, Eugen Rogozinnikov, and Anna Wienhard. Symmetric Spaces for Groups over Involutive Algebras and Applications to Higgs Bundles. *CoRR*, abs/2509.02002, 2025. URL https://arxiv.org/abs/2509.02002.

Zhiwu Huang and Luc Van Gool. A Riemannian Network for SPD Matrix Learning. In *AAAI*, pp. 2036–2042, 2017.

Zhiwu Huang, Chengde Wan, Thomas Probst, and Luc Van Gool. Deep Learning on Lie Groups for Skeleton-Based Action Recognition. In *CVPR*, pp. 6099–6108, 2017.

Zhiwu Huang, Jiqing Wu, and Luc Van Gool. Building Deep Networks on Grassmann Manifolds. In *AAAI*, pp. 3279–3286, 2018.

Michael Kapovich, Bernhard Leeb, and John J. Millson. The Generalized Triangle Inequalities in Symmetric Spaces and Buildings with Applications to Algebra. *CoRR*, abs/math/0210256, 2005. URL https://arxiv.org/abs/math/0210256.

Michael Kapovich, Bernhard Leeb, and Joan Porti. Anosov Subgroups: Dynamical and Geometric Characterizations. *CoRR*, abs/1703.01647, 2017. URL https://arxiv.org/abs/1703.01647.

Reinmar J. Kobler, Jun ichiro Hirayama, Qibin Zhao, and Motoaki Kawanabe. SPD Domain-specific Batch Normalization to Crack Interpretable Unsupervised Domain Adaptation in EEG. In *NeurIPS*, pp. 6219–6235, 2022.

Guy Lebanon and John Lafferty. Hyperplane Margin Classifiers on the Multinomial Manifold. In *ICML*, pp. 66, 2004.

Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C. Kot. NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. *TPAMI*, 42(10): 2684–2701, 2019.

Federico López, Beatrice Pozzetti, Steve Trettel, Michael Strube, and Anna Wienhard. Symmetric Spaces for Graph Embeddings: A Finsler-Riemannian Approach. In *ICML*, pp. 7090–7101, 2021.

Galileo Namata, Ben London, Lise Getoor, and Bert Huang. Query-driven Active Surveying for Collective Classification. In *10th International Workshop on Mining and Learning with Graphs*, volume 8, pp. 1, 2012a.

Galileo Namata, Ben London, Lise Getoor, and Bert Huang. Query-driven Active Surveying for Collective Classification. In *Workshop on Mining and Learning with Graphs*, 2012b.

Xuan Son Nguyen. The Gyro-Structure of Some Matrix Manifolds. In *NeurIPS*, pp. 26618–26630, 2022.

Xuan Son Nguyen and Nistor Grozavu. Batch normalization for neural networks on complex domains. *CoRR*, abs/2605.00467, 2026. URL https://arxiv.org/abs/2605.00467.

Xuan Son Nguyen and Shuo Yang. Building Neural Networks on Matrix Manifolds: A Gyrovector Space Approach. In *ICML*, pp. 26031–26062, 2023.

Xuan Son Nguyen, Shuo Yang, and Aymeric Histace. Matrix Manifold Neural Networks++. In *ICLR*, 2024.

Xuan Son Nguyen, Aymeric Histace, and Nistor Grozavu. Siegel Neural Networks. In *NeurIPS*, 2025.

Frank Nielsen. The SiegelKlein Disk: Hilbert Geometry of the Siegel Disk Domain. *Entropy*, 22 (9), 2020.

Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. MAtt: A Manifold Attention Network for EEG Decoding. *CoRR*, abs/2210.01986, 2022. URL https://arxiv.org/abs/2210.01986.

Xavier Pennec. *Statistical Computing on Manifolds for Computational Anatomy*. Habilitation à diriger des recherches, Université Nice Sophia-Antipolis, 2006.

Salem Said, Hatem Hajri, Lionel Bombrun, and Baba C. Vemuri. Gaussian Distributions on Riemannian Symmetric Spaces: Statistical Learning With Structured Covariance Matrices. *IEEE Transactions on Information Theory*, 64(2):752–772, 2018.

Prithviraj Sen, Galileo Mark Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, and Tina Eliassi-Rad. Collective Classification in Network Data. *AI Magazine*, 29(3):93–106, 2008.

Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis. In *CVPR*, pp. 1010–1019, 2016.

Carl Ludwig Siegel. Symplectic Geometry. *American Journal of Mathematics*, 65:1–86, 1943.

Max van Spengler, Erwin Berkhout, and Pascal Mettes. Poincaré ResNet. In *ICCV*, pp. 5396–5405, 2023.

K. Yun, J. Honorio, D. Chattopadhyay, T. L. Berg, and D. Samaras. Two-person Interaction Detection Using Body-pose Features And Multiple Instance Learning. In *CVPRW*, pp. 28–35, 2012.

Muhan Zhang and Yixin Chen. Link Prediction Based on Graph Neural Networks. *CoRR*, abs/1802.09691, 2018. URL https://arxiv.org/abs/1802.09691.

Wei Zhao, Federico Lopez, J. Maxwell Riestenberg, Michael Strube, Diaaeldin Taha, and Steve Trettel. Modeling Graphs Beyond Hyperbolic: Graph Neural Networks in Symmetric Positive Definite Matrices. *CoRR*, abs/2306.14064, 2023. URL https://arxiv.org/abs/2306.14064.

A Experiments

A.1 Radar Clutter Classification

A.1.1 Datasets and Experimental Settings

The AR model is given by

$$u_t + \sum_{l=1}^{r} c_l u_{t-l} = v_t,$$

where r ($r > 1$) is the order of the AR model, $u_t \in \mathbb{C}_n$ is the vector of signals at time t, $c_l \in \mathbb{C}^{n\times n}, l = 1, \ldots, r$ are the prediction coefficients (AR parameters), and $v_t \in \mathbb{C}_n$ is the prediction error at time t which is assumed to be a multidimensional Gaussian random variable.

The statistics of the datasets are given in Tab. 6.

A.1.2 Optimization and Hyperparameters

Similar to Nguyen & Grozavu (2026), we parameterize the Fréchet mean $\mathbf{X}$ of a set of points $\mathbf{X}_1, \ldots, \mathbf{X}_h \in \mathbb{HB}_n$ by using parameters on Euclidean space. Let $\mathbf{Y} = \varphi^{(-1)}(\mathbf{X}) = \mathbf{Y}_0 + \mathbf{Y}_1\mathbf{i} \in \mathbb{HU}_n$, where $\mathbf{Y}_0 = \frac{1}{2}(\mathbf{Y} + \mathbf{Y}^H)$ and $\mathbf{Y}_1 = \frac{1}{2\mathbf{i}}(\mathbf{Y} - \mathbf{Y}^H) \in \mathbb{H}_n^+$. We parameterize matrix $\mathbf{Y}_0$ by two vectors $\mathbf{u}_0 \in \mathbb{R}^{\frac{n(n+1)}{2}}$ and $\mathbf{v}_0 \in \mathbb{R}^{\frac{n(n-1)}{2}}$ such that $\mathbf{Y}_0 = (f_1(\mathbf{u}_0) + f_1(\mathbf{u}_0)^T) + (f_2(\mathbf{v}_0) - f_2(\mathbf{v}_0)^T)\mathbf{i}$, where $f_1(\mathbf{u}_0)$ transforms vector $\mathbf{u}_0$ into a $n \times n$ lower triangular matrix, and $f_2(\mathbf{v}_0)$ transforms vector $\mathbf{v}_0$ into a $n \times n$ strictly lower triangular matrix. Similarly, matrix $\mathbf{Y}_1$ is parameterized by two vectors $\mathbf{u}_1 \in \mathbb{R}^{\frac{n(n+1)}{2}}$ and $\mathbf{v}_1 \in \mathbb{R}^{\frac{n(n-1)}{2}}$ such that $\mathbf{Y}_1 = (f_1(\mathbf{u}_1) + f_1(\mathbf{u}_1)^T) + (f_2(\mathbf{v}_1) - f_2(\mathbf{v}_1)^T)\mathbf{i}$. We use the eigenvalue decomposition of $\mathbf{Y}_1$ to project it to $\mathbb{H}_n^+$. The Fréchet mean $\mathbf{X}$ is obtained by solving the following optimization problem

$$(\mathbf{u}_0, \mathbf{u}_1, \mathbf{v}_0, \mathbf{v}_1) = \underset{\mathbf{u}_0, \mathbf{u}_1 \in \mathbb{R}^{\frac{n(n+1)}{2}}, \mathbf{v}_0, \mathbf{v}_1 \in \mathbb{R}^{\frac{n(n-1)}{2}}}{\arg\min} \sum_{l=1}^{h} d^2_{\mathbb{HB}_n}(\mathbf{X}_l, f_1(\mathbf{u}_0) + f_1(\mathbf{u}_0)^T - f_2(\mathbf{v}_1) + f_2(\mathbf{v}_1)^T + (f_1(\mathbf{u}_1) + f_1(\mathbf{u}_1)^T + f_2(\mathbf{v}_0) - f_2(\mathbf{v}_0)^T)\mathbf{i}),$$

where $d_{\mathbb{HB}_n}(\cdot)$ is the distance function defined in Appendix C.1.

Our networks are implemented in the Pytorch framework and trained using cross-entropy loss and Adadelta optimizer for 2000 epochs. The learning rate and the batch size are set to $1e-2$ and 25, respectively. The number of iterations for computing the Fréchet mean is set to 5. Results are averaged over 5 random parameter initializations for each model. Our experiments are performed using a machine with Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz.

Table 6: Statistics of the datasets for the radar clutter classification experiments.

Parameter	Dimension	Length	Number of classes	Order	Size
Dataset 1	30	50	20	3	950
Dataset 2	30	50	40	2	450
Dataset 3	30	50	60	3	650
Dataset 4	30	50	80	4	900
Dataset 5	30	50	100	2	1000
Dataset 6	30	50	120	2	1200

Table 7: Numbers of parameters (MB) for the Siegel model and our model in the radar clutter classification experiments.

Model	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
SiegelNetBN (Nguyen & Grozavu, 2026)	90951	180971	270991	361011	451031	541051
HBNetBN (Ours)	163821	325841	487861	649881	811901	973921

A.1.3 More Results

Tabs. 7 and 8 report the numbers of parameters and the training times of the Siegel model and our model in the radar clutter classification experiments.

A.2 Action Recognition

A.2.1 Datasets and Experimental Settings

SBU Interaction It is an interaction action dataset which contains 282 sequences in 8 action classes created from 7 subjects. Each action is performed by two subjects where each subject has 15 joints.

NTU60 It has 56880 sequences of 3D skeleton data with 60 classes. Each frame contains the 3D coordinates of 25 or 50 body joints. We use the mutual (interaction) actions for classification (11 classes).

NTU120 It has 114480 sequences in 120 action classes, captured by 106 subjects with three cameras views. Each frame contains the 3D coordinates of 25 or 50 body joints. We use the mutual (interaction) actions for classification (26 classes). For the cross-subject (XSubject120) protocol (Liu et al., 2019), the 106 subjects are split into training and testing groups where each group contains 53 subjects. For the cross-setup (XSetup120) protocol (Liu et al., 2019), training data contain samples with even setup IDs, and testing data contain samples with odd setup IDs.

A.2.2 Optimization and Hyperparameters

The baselines have the same architectures proposed in the original papers (Brooks et al., 2019; Chen et al., 2024). Our networks are implemented in the Pytorch framework and trained using cross-entropy loss and Adadelta optimizer for 5000 epochs. The learning rate and the batch size are set to $1e-2$ and 32, respectively. Results are computed over 5 random parameter initializations for each model. Our experiments are performed using a machine with Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz.

A.2.3 More Results

Tab. 9 shows the results in Tab. 2 along with those of GyroSpd++ (Nguyen et al., 2024).

Table 8: Training times (seconds) per batch of the Siegel model and our model in the radar clutter classification experiments (Hardware: Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz)

Model	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
SiegelNetBN (Nguyen & Grozavu, 2026)	0.23	0.22	0.21	0.18	0.18	0.20
HBNetBN (Ours)	0.36	0.35	0.33	0.31	0.33	0.32

Table 9: Action recognition results (mean accuracy ± standard deviation) across five runs. XSubject corresponds to the cross-subject setting of NTU60 and NTU120. XView and XSetup correspond to the cross-view and cross-setup settings of NTU60 and NTU120, respectively. For NTU60, we use the first 20 training/testing sequences of each action class. For NTU120, we use the first 10 training/testing sequences of each action class.

Method	SBU		NTU60		NTU120	
	Split 1	Split 2	XSubject	XView	XSubject	XSetup
SPDNetBN (Brooks et al., 2019)	72.53±1.31	64.69±0.29	44.41±1.10	54.33±1.59	27.89±0.29	32.54±0.79
LieBN-AIM-(1.5) (Chen et al., 2024)	72.40±1.27	63.36±1.41	43.41±0.92	53.33±1.46	28.32±1.43	33.13±0.53
LieBN-LCM-(0.5) (Chen et al., 2024)	69.66±4.09	58.54±2.13	41.91±0.89	52.24±1.13	25.09±1.12	32.07±1.42
GyroSpd++ (Nguyen et al., 2024)	72.84±1.22	64.25±0.26	42.25±0.94	53.08±1.34	26.52±0.46	32.43±0.80
SiegelNetBN (Nguyen & Grozavu, 2026)	**76.80±0.68**	61.61±0.75	44.75±1.09	55.26±0.80	30.60±0.45	33.26±0.62
HBNetBN (Ours)	75.52±0.56	**68.12±1.06**	**46.28±0.70**	**57.01±0.58**	**32.63±0.65**	**34.02±0.59**

A.3 Graph Reconstruction

A.3.1 Datasets and Experimental Settings

We generate datasets of graphs with different geometric structures and sizes. The statistics of the datasets are summarized in Tab. 10.

A.3.2 Optimization and Hyperparameters

We create graph node embeddings on a space $\mathcal{M}$ by optimizing a distance-based loss function (Gu et al., 2019; López et al., 2021). Given the distances $\{d_G(l_1, l_2)\}_{l_1,l_2=1}^{N}$ between all pairs of connected nodes l_1 and l_2, the loss function is computed as:

$$\mathcal{L}(\mathbf{X}) = \sum_{l_1,l_2=1}^{N} \left| \left(\frac{d_{\mathcal{M}}(\mathbf{X}_{l_1}, \mathbf{X}_{l_2})}{d_G(l_1, l_2)} \right)^2 - 1 \right|,$$

where $d_{\mathcal{M}}(\cdot, \cdot)$ is the distance function on $\mathcal{M}$, $\mathbf{X}_{l_1}$ and $\mathbf{X}_{l_2}$ are the embeddings of nodes l_1 and l_2, respectively. The input distances are the shortest distances between pairs of connected nodes.

Our evaluation metric is the average of distortions over all pairs of points (López et al., 2021). The distortion of a pair of connected nodes l_1 and l_2 is given as

$$\text{distortion}(l_1, l_2) = \frac{\|d_{\mathcal{M}}(\mathbf{X}_{l_1}, \mathbf{X}_{l_2}) - d_G(l_1, l_2)\|}{d_G(l_1, l_2)}.$$

The distance function on $\mathbb{HB}_n$ is computed as in Algorithm 1. The node embeddings on $\mathbb{HB}_n$ are converted from those on $\mathbb{HU}_n$ which are parameterized using the same method in Section A.1.2.

Our experiments are based on the official code of López et al. (2021)[1]. The networks are implemented in the Pytorch framework and trained for a maximum of 3000 epochs. The batch size and the learning rate are set to 2048 and $1e-3$, respectively. The learning rate is reduced by a factor of five if the model does not improve the performance after 50 epochs (López et al., 2021). We use Riemannian SGD optimizer for the SPD and Siegel models, and SGD optimizer for our model. Early

[1] https://github.com/fedelopez77/sympa

Table 10: Statistics of the datasets for graph reconstruction experiments.

Graph	Nodes	Edges	Triples	Grid Layout	Tree Branching	Tree Height
Grid	243	810	29403	5×125		
Tree	40	39	780		3	3
Tree $\times$ Grid	1080	3213	582660	3×27	3	3
Grid $\times$ Tree	1080	3213	582660	3×27	3	3
Tree $\diamond$ Grid	364	363	66066	4×4	3	5
Grid $\diamond$ Tree	63	62	1953	4×4	2	5

Table 11: Statistics of the datasets for the node classification experiments.

Dataset	#Nodes	#Edges	#Classes	#Features
Disease	1044	1043	2	1000
Airport	3188	18631	4	4
Pubmed	19717	44338	3	500
Cora	2708	5429	7	1433

stopping based on the average distortion is used when the model does not improve after 150 epochs. Results are computed over five random parameter initializations for each model. Our experiments are performed using a machine with Intel(R) Xeon(R) Gold 6230R CPU @ 2.10GHz.

A.4 Node Classification

A.4.1 Datasets and Experimental Settings

Airport (Chami et al., 2019) It is a flight network dataset from OpenFlights.org where nodes represent airports, edges represent the airline Routes, and node labels are the populations of the country where the airport belongs.

Pubmed (Namata et al., 2012b) It is a standard benchmark describing citation networks where nodes represent scientific papers in the area of medicine, edges are citations between them, and node labels are academic (sub)areas.

Cora (Sen et al., 2008) It is a citation network where nodes represent scientific papers in the area of machine learning, edges are citations between them, and node labels are academic (sub)areas. Each publication in the dataset is described by a 0/1-valued word vector indicating the absence/presence of the corresponding word from the dictionary.

Disease (Anderson & May, 1992) It models the spread of infectious disease on a graph using the SusceptibleInfectedRecovered (SIR) epidemic model. Each node represents an individual, while edges represent disease transmission pathways between individuals. The class of a node indicates whether it is infected or uninfected individual.

We use the train, validation, and test splits as in Zhao et al. (2023). The statistics of the datasets are summarized in Tab. 11.

A.4.2 Optimization and Hyperparameters

The BiMap layer of our network reduces the dimensions of the input matrices by a factor of two. The isometry map used for the feature transform operation of our network is given (Zhao et al., 2023) by

$$\mathbf{M} \odot \mathbf{P} = \mathbf{M}\mathbf{P}\mathbf{M}^T,$$

where $\mathbf{M}$ is an invertible real matrix, and $\mathbf{P}$ is an SPD matrix.

Table 12: Numbers of parameters (MB) for the SPD model and our model in the node classification experiments. The SPD model and our model learn the node embeddings on Sym^+_{14} and $\mathbb{HB}_8$, respectively.

Model	Airport	Pubmed	Cora	Disease
SPD-GCN (Zhao et al., 2023)	3392	99690	282982	197584
HB-GCN (Ours)	3044	34651	94911	66514

Our experiments are based on the official code of Zhao et al. (2023)[2]. The networks are implemented in the Pytorch framework and trained using cross-entropy loss and Adam optimizer for a maximum of 500 epochs. The batch size is set to the total number of graph nodes in a dataset (Chami et al., 2019; Zhao et al., 2023). The learning rate is set to $1e-2$. Early stopping is used when the loss on the validation set has not decreased for 200 epochs. Each network has two layers that perform message passing twice at one iteration (Zhao et al., 2023). Our experiments are performed using a machine with Intel(R) Xeon(R) Silver 4510.

A.4.3 More Results

Tab. 12 provides the numbers of parameters for the SPD model and our model in the node classification experiments.

B Complexity Analysis

Below, we provide a time complexity analysis of some components of the proposed method.

- The computation of the automorphisms and their inverses defined in Eqs. (3) and (4) can be performed in $O(n^3)$ time.
- The distance function in Algorithm 1 and the one in Eq. (10) can be computed in $O(n^3)$ time.

C Additional Details of the Proposed Layers

C.1 BN Layers

Our BN algorithm is shown in Algorithm 2. The function $\mathfrak{M}_{\mathbb{HB}_n}(\mathbf{X}_1,\ldots,\mathbf{X}_h)$ computes the Fréchet mean of the set of h points $\mathbf{X}_1,\ldots,\mathbf{X}_h \in \mathbb{HB}_n$ defined as

$$\mathfrak{M}_{\mathbb{HB}_n}(\mathbf{X}_1,\ldots,\mathbf{X}_h) = \underset{\mathbf{X}\in\mathbb{HB}_n}{\arg\min} \sum_{l=1}^{h} d^2_{\mathbb{HB}_n}(\mathbf{X}_l,\mathbf{X}),$$

where $d_{\mathbb{HB}_n}(\cdot,\cdot)$ is the distance function on $\mathbb{HB}_n$ given by

$$d^2_{\mathbb{HB}_n}(\mathbf{X},\mathbf{Y}) = \mathrm{Tr}\left(\log^2\left((\mathbf{I}+\mathbf{C}^{\frac{1}{2}})(\mathbf{I}-\mathbf{C}^{\frac{1}{2}})^{-1}\right)\right),$$

where $\mathbf{X},\mathbf{Y}\in\mathbb{HB}_n$, $\mathrm{Tr}(\cdot)$ denotes the trace operator, $\log(\cdot)$ denotes the matrix logarithm, and $\mathbf{C}$ is computed as

$$\mathbf{C} = (\mathbf{Y}-\mathbf{X})(\mathbf{I}-\mathbf{X}^H\mathbf{Y})^{-1}(\mathbf{Y}^H-\mathbf{X}^H)(\mathbf{I}-\mathbf{X}\mathbf{Y}^H)^{-1}.$$

We use the following Hermitian form

$$d^2_{\mathbb{HB}_n}(\mathbf{X},\mathbf{Y}) = \left\|\log\left((\mathbf{I}+\mathbf{C}^{\frac{1}{2}})(\mathbf{I}-\mathbf{C}^{\frac{1}{2}})^{-1}\right)\right\|^2_F, \tag{10}$$

where $\mathbf{C}$ is computed as

$$\mathbf{C} = \mathbf{I} - (\mathbf{I}-\mathbf{X}\mathbf{X}^H)^{\frac{1}{2}}(\mathbf{I}-\mathbf{Y}\mathbf{X}^H)^{-1}(\mathbf{I}-\mathbf{Y}\mathbf{Y}^H)(\mathbf{I}-\mathbf{X}\mathbf{Y}^H)^{-1}(\mathbf{I}-\mathbf{X}\mathbf{X}^H)^{\frac{1}{2}}.$$

[2] https://github.com/andyweizhao/SPD4GNNs

Algorithm 2 BN on $\mathbb{HB}_n$

Training Phase
Input: Batch of h points $\{\mathbf{X}_l\}_{l=1}^h$ on $\mathbb{HB}_n$, running mean $\mathbf{M}_r$, bias $\mathbf{G}$, momentum η.
$\mathbf{M}_b \leftarrow \mathfrak{M}_{\mathbb{HB}_n}(\mathbf{X}_1, \ldots, \mathbf{X}_h)$
$\mathbf{M}_r \leftarrow \gamma_{\mathbf{M}_r, \mathbf{M}_b}(\eta)$
for $l = 1$ **to** h **do**
  $\overline{\mathbf{X}}_l \leftarrow \phi_{\mathbf{M}_b}(\mathbf{X}_l)$
  $\tilde{\mathbf{X}}_l \leftarrow \phi_{\mathbf{G}}^{(-1)}(\overline{\mathbf{X}}_l)$
end for
Output: Normalized batch $\{\tilde{\mathbf{X}}_l\}_{l=1}^h$.
Testing Phase
Input: Batch of h points $\{\mathbf{X}_l\}_{l=1}^h$ on $\mathbb{HB}_n$, running mean $\mathbf{M}_r$, learned bias $\mathbf{G}$.
for $l = 1$ **to** h **do**
  $\overline{\mathbf{X}}_l \leftarrow \phi_{\mathbf{M}_r}(\mathbf{X}_l)$
  $\tilde{\mathbf{X}}_l \leftarrow \phi_{\mathbf{G}}^{(-1)}(\overline{\mathbf{X}}_l)$
end for
Output: Normalized batch $\{\tilde{\mathbf{X}}_l\}_{l=1}^h$.

The functions $\phi_{\mathbf{M}_b}(\cdot)$ ($\phi_{\mathbf{M}_r}(\cdot)$) and $\phi_{\mathbf{G}}^{(-1)}(\cdot)$ are the automorphism and the automorphism inverse given in Eqs. (3) and (4), respectively. The function $\gamma_{\mathbf{M}_r, \mathbf{M}_b}(\eta)$ computes the almost geodesic joining $\mathbf{M}_r$ and $\mathbf{M}_b$ and is given by

$$\gamma_{\mathbf{M}_r, \mathbf{M}_b}(\eta) = \phi_{\mathbf{M}_r}^{(-1)}\left(\frac{1}{\|\phi_{\mathbf{M}_r}(\mathbf{M}_b)\|_2}\frac{(1+\|\phi_{\mathbf{M}_r}(\mathbf{M}_b)\|_2)^\eta - (1-\|\phi_{\mathbf{M}_r}(\mathbf{M}_b)\|_2)^\eta}{(1+\|\phi_{\mathbf{M}_r}(\mathbf{M}_b)\|_2)^\eta + (1-\|\phi_{\mathbf{M}_r}(\mathbf{M}_b)\|_2)^\eta}\phi_{\mathbf{M}_r}(\mathbf{M}_b)\right),$$

where $\|\cdot\|_2$ denotes the spectral norm of a matrix given by its largest singular value.

C.2 Point-to-hyperplane Distances on HPD Manifolds

The approaches in Lebanon & Lafferty (2004); Ganea et al. (2018); Nguyen (2022); Nguyen & Yang (2023) can be seamlessly extended to HPD manifolds. To make the paper self-contained, in this section, we present the definitions and results needed for our derivations in the main paper.

Definition C.1 (**HPD Binary Operation**). *For $\mathbf{P}, \mathbf{Q} \in \mathbb{H}_n^+$, the binary operation $\oplus$ is defined as*

$$\mathbf{P} \oplus \mathbf{Q} = \mathbf{P}^{\frac{1}{2}}\mathbf{Q}\mathbf{P}^{\frac{1}{2}}.$$

Definition C.2 (**HPD Inverse Operation**). *For $\mathbf{P} \in \mathbb{H}_n^+$, the inverse operation $\ominus$ is defined as*

$$\ominus\mathbf{P} = \mathbf{P}^{-1}.$$

Definition C.3 (**HPD Scalar Multiplication**). *For $\mathbf{P} \in \mathbb{H}_n^+$, $t \in \mathbb{R}$, the scalar multiplication $\otimes$ is defined as*

$$t \otimes \mathbf{P} = \mathbf{P}^t.$$

Definition C.4 (**HPD Inner Product**). *Let $\mathbf{P}, \mathbf{Q} \in \mathbb{H}_n^+$. Then the HPD inner product of $\mathbf{P}$ and $\mathbf{Q}$ is defined as*

$$\langle \mathbf{P}, \mathbf{Q} \rangle = \langle \mathrm{Log}_{\mathbf{I}}(\mathbf{P}), \mathrm{Log}_{\mathbf{I}}(\mathbf{Q}) \rangle_{\mathbf{I}},$$

where $\langle \cdot, \cdot \rangle_{\mathbf{I}}$ denotes the inner product at $\mathbf{I}$ given by the Riemannian metric on $\mathbb{H}_n^+$ (Said et al., 2018).

Definition C.5 (**HPD Hypergyroplanes**). *For $\mathbf{P} \in \mathbb{H}_n^+$, $\mathbf{W} \in T_{\mathbf{P}}\mathbb{H}_n^+$, HPD hypergyroplanes are defined as*

$$\mathcal{H}_{\mathbf{W},\mathbf{P}} = \{\mathbf{Q} \in \mathbb{H}_n^+ : \langle \mathrm{Log}_{\mathbf{P}}(\mathbf{Q}), \mathbf{W} \rangle_{\mathbf{P}} = 0\},$$

where $\mathrm{Log}_{\mathbf{P}}(\mathbf{Q})$ denotes the logarithmic map of $\mathbf{Q}$ at $\mathbf{P}$, and $\langle \cdot, \cdot \rangle_{\mathbf{P}}$ denotes the inner product at $\mathbf{P}$ given by the Riemannian metric on $\mathbb{H}_n^+$.

We first extend the notion of gyroangles on gyrovector spaces to HPD manifolds.

Definition C.6 (**HPD Gyrocosine Function and HPD Gyroangles).** *Let* $\mathbf{P}, \mathbf{Q}$, *and* $\mathbf{R}$ *be three distinct HPD gyropoints (HPD matrices) in a gyrovector space* $(\mathbb{H}_n^+, \oplus, \otimes)$. *The HPD gyrocosine of the measure of the HPD gyroangle* α, $0 \leq \alpha \leq \pi$, *between* $\ominus\mathbf{P} \oplus \mathbf{Q}$ *and* $\ominus\mathbf{P} \oplus \mathbf{R}$ *is given by the equation*

$$\cos\alpha = \frac{\langle \ominus\mathbf{P} \oplus \mathbf{Q}, \ominus\mathbf{P} \oplus \mathbf{R}\rangle}{\| \ominus \mathbf{P} \oplus \mathbf{Q}\| \| \ominus \mathbf{P} \oplus \mathbf{R}\|},$$

where $\|\cdot\|$ *denotes the HPD norm induced by the HPD inner product. The HPD gyroangle* α *is denoted by* $\alpha = \angle\mathbf{QPR}$.

We can now define the margin distance from an HPD matrix to an HPD hypergyroplane.

Definition C.7 (**HPD Pseudo-gyrodistance from an HPD Matrix to an HPD Hypergyroplane).** *Let* $\mathcal{H}_{\mathbf{W},\mathbf{P}}$ *be an HPD hypergyroplane. Then the HPD pseudo-gyrodistance from any* $\mathbf{X} \in \mathbb{H}_n^+$ *to* $\mathcal{H}_{\mathbf{W},\mathbf{P}}$ *is defined as*

$$\bar{d}(\mathbf{X}, \mathcal{H}_{\mathbf{W},\mathbf{P}}) = \sin(\angle\mathbf{XP}\bar{\mathbf{Q}})d(\mathbf{X}, \mathbf{P}),$$

where $\bar{\mathbf{Q}}$ *is given by*

$$\bar{\mathbf{Q}} = \underset{\mathbf{Q}\in\mathcal{H}_{\mathbf{W},\mathbf{P}}\setminus\{\mathbf{P}\}}{\arg\max} \Big(\frac{\langle \ominus_g\mathbf{P} \oplus_g \mathbf{Q}, \ominus_g\mathbf{P} \oplus_g \mathbf{X}\rangle}{\| \ominus_g \mathbf{P} \oplus_g \mathbf{Q}\|.\| \ominus_g \mathbf{P} \oplus_g \mathbf{X}\|}\Big).$$

By convention, $\sin(\angle\mathbf{XPQ}) = 0$ *for any* $\mathbf{X}, \mathbf{Q} \in \mathcal{H}_{\mathbf{W},\mathbf{P}}$.

We have the following definition.

Definition C.8. *Let* $\mathbf{P} \in \mathbb{H}_n^+$, $\mathbf{W} \in T_{\mathbf{P}}\mathbb{H}_n^+$, *and let* $\mathcal{H}_{\mathbf{W},\mathbf{P}}$ *be an HPD hypergyroplane. Then the HPD pseudo-gyrodistance from any* $\mathbf{X} \in \mathbb{H}_n^+$ *to* $\mathcal{H}_{\mathbf{W},\mathbf{P}}$ *is given by*

$$\bar{d}(\mathbf{X}, \mathcal{H}_{\mathbf{W},\mathbf{P}}) = \frac{|\langle \log(\mathbf{P}^{-\frac{1}{2}}\mathbf{X}\mathbf{P}^{-\frac{1}{2}}), \mathbf{P}^{-\frac{1}{2}}\mathbf{W}\mathbf{P}^{-\frac{1}{2}}\rangle_F|}{\|\mathbf{P}^{-\frac{1}{2}}\mathbf{W}\mathbf{P}^{-\frac{1}{2}}\|_F},$$

where $\log(\cdot)$ *denotes the matrix logarithm,* $|\cdot|$ *denotes the absolute value,* $\langle\cdot,\cdot\rangle_F$ *and* $\|\cdot\|_F$ *denote the Frobenius inner product and Frobenius norm, respectively,*

C.3 Extension of the Proposed Classification Layer to the Product Space Setting

The proposed classification layer can be extended to the product space setting as follows.

Definition C.9. *Let* $\mathbf{Z} = (\mathbf{Y}, \mathbf{Z}^1, \ldots, \mathbf{Z}^{r-1}) \in \mathrm{Sym}_n^+ \times (\mathbb{HB}_n^+)^{r-1}$ *be an input of the classification layer. Let* $\mathbf{Z}_1^j, \mathbf{Z}_2^j \in \mathrm{Mat}_n(\mathbb{R})$ *be the real and imaginary parts of* $\varphi^{(-1)}(\mathbf{Z}^j) \in \mathbb{HU}_n, j = 1, \ldots, r-1$. *Let* $\tilde{\mathbf{Z}}_1^j = \frac{1}{2}\Big((\mathbf{Z}_1^j + (\mathbf{Z}_1^j)^T) + (\mathbf{Z}_2^j - (\mathbf{Z}_2^j)^T)\mathbf{i}\Big)$ *and* $\tilde{\mathbf{Z}}_2^j = \frac{1}{2}\Big((\mathbf{Z}_2^j + (\mathbf{Z}_2^j)^T) + ((\mathbf{Z}_1^j)^T - \mathbf{Z}_1^j)\mathbf{i}\Big)$. *Define*

$$\mathbf{X}_j = \begin{pmatrix} \tilde{\mathbf{Z}}_2^j + \tilde{\mathbf{Z}}_1^j(\tilde{\mathbf{Z}}_2^j)^{-1}(\tilde{\mathbf{Z}}_1^j)^H & \tilde{\mathbf{Z}}_1^j(\tilde{\mathbf{Z}}_2^j)^{-1} \\ (\tilde{\mathbf{Z}}_2^j)^{-1}(\tilde{\mathbf{Z}}_1^j)^H & (\tilde{\mathbf{Z}}_2^j)^{-1} \end{pmatrix}$$

for $j = 1, \ldots, r-1$. *Then the probability of class* l *is computed as*

$$p(y = l|\mathbf{Z}) \propto \exp\left(\frac{\Big|\langle \log(\mathbf{P}_0^{-\frac{1}{2}}\mathbf{Y}\mathbf{P}_0^{-\frac{1}{2}}), \mathbf{P}_0^{-\frac{1}{2}}\mathbf{W}_0\mathbf{P}_0^{-\frac{1}{2}}\rangle_F + \sum_{j=1}^{r-1}\langle \log(\mathbf{P}_j^{-\frac{1}{2}}\mathbf{X}_j\mathbf{P}_j^{-\frac{1}{2}}), \mathbf{P}_j^{-\frac{1}{2}}\mathbf{W}_j\mathbf{P}_j^{-\frac{1}{2}}\rangle_F\Big|}{\sqrt{\Big\|\mathbf{P}_0^{-\frac{1}{2}}\mathbf{W}_0\mathbf{P}_0^{-\frac{1}{2}}\Big\|_F^2 + \sum_{j=1}^{r-1}\Big\|\mathbf{P}_j^{-\frac{1}{2}}\mathbf{W}_j\mathbf{P}_j^{-\frac{1}{2}}\Big\|_F^2}}\right),$$

where $\mathbf{P}_0 \in \mathrm{Sym}_n^+$, $\mathbf{W}_0 \in T_{\mathbf{P}_0}\mathrm{Sym}_n^+$, $\mathbf{P}_j \in \mathbb{H}_n^+$, $\mathbf{W}_j \in T_{\mathbf{P}_j}\mathbb{H}_n^+, j = 1, \ldots, r-1$ *are the parameters of the layer.*

D Limitation of Our Method

Our networks are based on the computation of automorphisms on the studied domains which involves matrix inversion and SVD operation. Therefore, they suffer from high computational cost when training with high-dimensional input.

Although our method has proven effective for some machine learning tasks, the framework developped so far only allows us to build a limited number of network building blocks, i.e., BN layers and classification layers. The proposed networks built upon these building blocks are mainly designed to tackle classification problems. To further evaluate the effectiveness of the proposed framework, other fundamental building blocks (e.g., attention layers) and deep architectures must be considered.

E MATHEMATICAL PROOFS

E.1 PROOF OF PROPOSITION 3.1

Proof. Let $\mathbf{Z}_0$ be a point of $\mathfrak{B}$, then

$$\bar{\sigma}_{\mathbb{C}}(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)(\mathbf{Z}_0+\mathbf{i}\mathbf{I}) = -\mathbf{i}(\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)+\mathbf{i}\mathbf{I})(\mathbf{Z}_0+\mathbf{i}\mathbf{I}),$$
$$\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0+\mathbf{i}\mathbf{I})(\mathbf{I}+\mathbf{i}\mathbf{Z}_0) = (\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)-\mathbf{i}\mathbf{I})(\mathbf{I}+\mathbf{i}\mathbf{Z}_0).$$

Hence

$$\bar{\sigma}_{\mathbb{C}}(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)(\mathbf{Z}_0+\mathbf{i}\mathbf{I}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0+\mathbf{i}\mathbf{I})(\mathbf{I}+\mathbf{i}\mathbf{Z}_0) = 2\mathbf{i}(\mathbf{I}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)\mathbf{Z}_0).$$

Let $\mathbf{R}$ be a solution of $(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)\mathbf{R}=\mathbf{0}$, then $\mathbf{R}=-\mathbf{i}\mathbf{Z}_0\mathbf{R}$ and we have

$$\begin{aligned}\bar{\sigma}_{\mathbb{C}}(\mathbf{R})(\mathbf{I}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)\mathbf{Z}_0)\mathbf{R} &= \bar{\sigma}_{\mathbb{C}}(\mathbf{R})\mathbf{R} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0\mathbf{R})\mathbf{Z}_0\mathbf{R}\\ &= \mathbf{i}\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0\mathbf{R})(-\mathbf{i}\mathbf{Z}_0\mathbf{R}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0\mathbf{R})\mathbf{Z}_0\mathbf{R}\\ &= \mathbf{0}.\end{aligned}$$

Since $\mathbf{I}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)\mathbf{Z}_0 \in (A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$, we deduce that $\mathbf{I}+\mathbf{i}\mathbf{Z}_0$ is invertible. Therefore

$$\begin{aligned}(\mathbf{Z}_0+\mathbf{i}\mathbf{I})(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)^{-1} - \left(\bar{\sigma}_{\mathbb{C}}(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)\right)^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0+\mathbf{i}\mathbf{I}) &= 2\mathbf{i}\left(\bar{\sigma}_{\mathbb{C}}(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)\right)^{-1}(\mathbf{I}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)\mathbf{Z}_0)(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)^{-1}\\ &= 2\mathbf{i}\bar{\sigma}_{\mathbb{C}}\left((\mathbf{I}+\mathbf{i}\mathbf{Z}_0)^{-1}\right)(\mathbf{I}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)\mathbf{Z}_0)(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)^{-1}.\end{aligned}$$

Let $\mathbf{Z}=(\mathbf{Z}_0+\mathbf{i}\mathbf{I})(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)^{-1}$, then we have proved that

$$\mathbf{Z}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}) = 2\mathbf{i}\bar{\sigma}_{\mathbb{C}}\left((\mathbf{I}+\mathbf{i}\mathbf{Z}_0)^{-1}\right)(\mathbf{I}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)\mathbf{Z}_0)(\mathbf{I}+\mathbf{i}\mathbf{Z}_0)^{-1},$$

which shows that $\mathbf{Z}\in\mathfrak{U}$.

Let $\mathbf{Z}$ be a point of $\mathfrak{U}$, then

$$\bar{\sigma}_{\mathbb{C}}(\mathbf{I}-\mathbf{i}\mathbf{Z})(\mathbf{I}-\mathbf{i}\mathbf{Z}) = (\bar{\sigma}_{\mathbb{C}}(\mathbf{Z})-\mathbf{i}\mathbf{I})(\mathbf{Z}+\mathbf{i}\mathbf{I}),$$
$$\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}-\mathbf{i}\mathbf{I})(\mathbf{Z}-\mathbf{i}\mathbf{I}) = (\bar{\sigma}_{\mathbb{C}}(\mathbf{Z})+\mathbf{i}\mathbf{I})(\mathbf{Z}-\mathbf{i}\mathbf{I}).$$

Hence

$$\bar{\sigma}_{\mathbb{C}}(\mathbf{I}-\mathbf{i}\mathbf{Z})(\mathbf{I}-\mathbf{i}\mathbf{Z}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}-\mathbf{i}\mathbf{I})(\mathbf{Z}-\mathbf{i}\mathbf{I}) = \frac{2}{\mathbf{i}}(\mathbf{Z}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z})).$$

Let $\mathbf{R}$ be a solution of $(\mathbf{I}-\mathbf{i}\mathbf{Z})\mathbf{R}=\mathbf{0}$, then $\mathbf{R}=\mathbf{i}\mathbf{Z}\mathbf{R}$ and we have

$$\begin{aligned}\bar{\sigma}_{\mathbb{C}}(\mathbf{R})(\mathbf{Z}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}))\mathbf{R} &= \bar{\sigma}_{\mathbb{C}}(\mathbf{R})\mathbf{Z}\mathbf{R} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}\mathbf{R})\mathbf{R}\\ &= -\mathbf{i}\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}\mathbf{R})\mathbf{Z}\mathbf{R} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}\mathbf{R})(\mathbf{i}\mathbf{Z}\mathbf{R})\\ &= -2\mathbf{i}\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}\mathbf{R})\mathbf{Z}\mathbf{R}.\end{aligned}$$

Since $\frac{1}{2\mathbf{i}}(\mathbf{Z}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z})) \in (A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$ and $\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}r)\mathbf{Z}r \in (A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_{\geq 0}$, we deduce that $\mathbf{I}-\mathbf{i}\mathbf{Z}$ is invertible. Therefore

$$\begin{aligned}\left(\bar{\sigma}_{\mathbb{C}}(\mathbf{I}-\mathbf{i}\mathbf{Z})\right)^{-1}\left(\bar{\sigma}_{\mathbb{C}}(\mathbf{I}-\mathbf{i}\mathbf{Z})(\mathbf{I}-\mathbf{i}\mathbf{Z}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}-\mathbf{i}\mathbf{I})(\mathbf{Z}-\mathbf{i}\mathbf{I})\right)(\mathbf{I}-\mathbf{i}\mathbf{Z})^{-1} =& \frac{2}{\mathbf{i}}\left(\bar{\sigma}_{\mathbb{C}}(\mathbf{I}-\mathbf{i}\mathbf{Z})\right)^{-1}\\ &(\mathbf{Z}-\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}))(\mathbf{I}-\mathbf{i}\mathbf{Z})^{-1}.\end{aligned}$$

Let $\mathbf{Z}_0 = (\mathbf{Z} - \mathbf{i}\mathbf{I})(\mathbf{I} - \mathbf{i}\mathbf{Z})^{-1}$, then

$$\begin{aligned}\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)\mathbf{Z}_0 &= \mathbf{I} - \left(\bar{\sigma}_{\mathbb{C}}(\mathbf{I} - \mathbf{i}\mathbf{Z})\right)^{-1} \bar{\sigma}_{\mathbb{C}}(\mathbf{Z} - \mathbf{i}\mathbf{I})(\mathbf{Z} - \mathbf{i}\mathbf{I})(\mathbf{I} - \mathbf{i}\mathbf{Z})^{-1} \\ &= \left(\bar{\sigma}_{\mathbb{C}}(\mathbf{I} - \mathbf{i}\mathbf{Z})\right)^{-1} \left(\bar{\sigma}_{\mathbb{C}}(\mathbf{I} - \mathbf{i}\mathbf{Z})(\mathbf{I} - \mathbf{i}\mathbf{Z}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z} - \mathbf{i}\mathbf{I})(\mathbf{Z} - \mathbf{i}\mathbf{I})\right)(\mathbf{I} - \mathbf{i}\mathbf{Z})^{-1}.\end{aligned}$$

We thus obtain

$$\begin{aligned}\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}_0)\mathbf{Z}_0 &= \frac{2}{\mathbf{i}} \left(\bar{\sigma}_{\mathbb{C}}(\mathbf{I} - \mathbf{i}\mathbf{Z})\right)^{-1} (\mathbf{Z} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}))(\mathbf{I} - \mathbf{i}\mathbf{Z})^{-1} \\ &= \frac{2}{\mathbf{i}} \bar{\sigma}_{\mathbb{C}}\left((\mathbf{I} - \mathbf{i}\mathbf{Z})^{-1}\right) (\mathbf{Z} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}))(\mathbf{I} - \mathbf{i}\mathbf{Z})^{-1},\end{aligned}$$

which shows that $\mathbf{Z}_0 \in \mathfrak{B}$.

□

E.2 Proof of Proposition 3.2

Proof. We prove the first assertion of the proposition. Let $\mathbf{M} = \begin{pmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{pmatrix} \in \mathrm{Sp}_2(A_{\mathbb{C}}, \bar{\sigma}_{\mathbb{C}})$, and let $\mathbf{Z}$ be any point in $\mathfrak{U}$, then

$$\begin{aligned}\mathbf{Z} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}) &= -\begin{pmatrix}\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}) & \mathbf{I}\end{pmatrix} \Omega \begin{pmatrix}\mathbf{Z} \\ \mathbf{I}\end{pmatrix} \\ &= -\begin{pmatrix}\bar{\sigma}_{\mathbb{C}}(\mathbf{Z}) & \mathbf{I}\end{pmatrix} \begin{pmatrix}\bar{\sigma}_{\mathbb{C}}(\mathbf{A}) & \bar{\sigma}_{\mathbb{C}}(\mathbf{C}) \\ \bar{\sigma}_{\mathbb{C}}(\mathbf{B}) & \bar{\sigma}_{\mathbb{C}}(\mathbf{D})\end{pmatrix} \Omega \begin{pmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{pmatrix} \begin{pmatrix}\mathbf{Z} \\ \mathbf{I}\end{pmatrix} \\ &= -\bar{\sigma}_{\mathbb{C}}\left(\begin{pmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{pmatrix} \begin{pmatrix}\mathbf{Z} \\ \mathbf{I}\end{pmatrix}\right) \Omega \begin{pmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{pmatrix} \begin{pmatrix}\mathbf{Z} \\ \mathbf{I}\end{pmatrix}\end{aligned}$$

Let $\mathbf{U} = \mathbf{A}\mathbf{Z} + \mathbf{B}$ and $V = \mathbf{C}\mathbf{Z} + \mathbf{D}$, then we have

$$\begin{aligned}\mathbf{Z} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}) &= -\bar{\sigma}_{\mathbb{C}}\begin{pmatrix}\mathbf{U} \\ \mathbf{V}\end{pmatrix}^{\mathsf{T}} \Omega \begin{pmatrix}\mathbf{U} \\ \mathbf{V}\end{pmatrix} \\ &= \bar{\sigma}_{\mathbb{C}}(\mathbf{V})\mathbf{U} - \bar{\sigma}_{\mathbb{C}}(\mathbf{U})\mathbf{V}.\end{aligned} \tag{11}$$

We thus get $\frac{1}{2\mathbf{i}}(\bar{\sigma}_{\mathbb{C}}(\mathbf{V})\mathbf{U} - \bar{\sigma}_{\mathbb{C}}(\mathbf{U})\mathbf{V}) = \frac{1}{2\mathbf{i}}(\mathbf{Z} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}))$ and therefore $\frac{1}{2\mathbf{i}}(\bar{\sigma}_{\mathbb{C}}(\mathbf{V})\mathbf{U} - \bar{\sigma}_{\mathbb{C}}(\mathbf{U})\mathbf{V}) \in (A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$.

Let $\mathbf{R}$ be a solution of $\mathbf{V}\mathbf{R} = \mathbf{0}$, then $\bar{\sigma}_{\mathbb{C}}(\mathbf{R})\bar{\sigma}_{\mathbb{C}}(\mathbf{V}) = \mathbf{0}$ and we have $\bar{\sigma}_{\mathbb{C}}(\mathbf{R})(\bar{\sigma}_{\mathbb{C}}(\mathbf{V})\mathbf{U} - \bar{\sigma}_{\mathbb{C}}(\mathbf{U})\mathbf{V})\mathbf{R} = \mathbf{0}$. Hence $\mathbf{V} \in A_{\mathbb{C}}^{\times}$. This proves the existence of $\mathbf{W} = (\mathbf{A}\mathbf{Z} + \mathbf{B})(\mathbf{C}\mathbf{Z} + \mathbf{D})^{-1}$. From Eq. (11),

$$\begin{aligned}\mathbf{Z} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}) &= -\bar{\sigma}_{\mathbb{C}}\left(\begin{pmatrix}(\mathbf{A}\mathbf{Z} + \mathbf{B})(\mathbf{C}\mathbf{Z} + \mathbf{D})^{-1} \\ \mathbf{I}\end{pmatrix} (\mathbf{C}\mathbf{Z} + \mathbf{D})\right) \Omega \begin{pmatrix}(\mathbf{A}\mathbf{Z} + \mathbf{B})(\mathbf{C}\mathbf{Z} + \mathbf{D})^{-1} \\ \mathbf{I}\end{pmatrix} (\mathbf{C}\mathbf{Z} + \mathbf{D}) \\ &= -\bar{\sigma}_{\mathbb{C}}\left(\begin{pmatrix}\mathbf{W} \\ \mathbf{I}\end{pmatrix} (\mathbf{C}\mathbf{Z} + \mathbf{D})\right) \Omega \begin{pmatrix}\mathbf{W} \\ \mathbf{I}\end{pmatrix} (\mathbf{C}\mathbf{Z} + \mathbf{D}) \\ &= -\bar{\sigma}_{\mathbb{C}}(\mathbf{C}\mathbf{Z} + \mathbf{D})\bar{\sigma}_{\mathbb{C}}\begin{pmatrix}\mathbf{W} \\ \mathbf{I}\end{pmatrix}^{\mathsf{T}} \Omega \begin{pmatrix}\mathbf{W} \\ \mathbf{I}\end{pmatrix} (\mathbf{C}\mathbf{Z} + \mathbf{D}) \\ &= \bar{\sigma}_{\mathbb{C}}(\mathbf{C}\mathbf{Z} + \mathbf{D})(\mathbf{W} - \bar{\sigma}_{\mathbb{C}}(\mathbf{W}))(\mathbf{C}\mathbf{Z} + \mathbf{D}),\end{aligned}$$

which leads to

$$\begin{aligned}\mathbf{W} - \bar{\sigma}_{\mathbb{C}}(\mathbf{W}) &= \bar{\sigma}_{\mathbb{C}}(\mathbf{C}\mathbf{Z} + \mathbf{D})^{-1}(\mathbf{Z} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}))(\mathbf{C}\mathbf{Z} + \mathbf{D})^{-1} \\ &= \bar{\sigma}_{\mathbb{C}}\left((\mathbf{C}\mathbf{Z} + \mathbf{D})^{-1}\right) (\mathbf{Z} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z}))(\mathbf{C}\mathbf{Z} + \mathbf{D})^{-1}.\end{aligned}$$

Since $(A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$ is closed by action of $A_{\mathbb{C}}^{\times}$, we obtain $\mathbf{W} - \bar{\sigma}_{\mathbb{C}}(\mathbf{W}) \in (A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$, thus $\mathbf{W} \in \mathfrak{U}$.

We now prove the second assertion of the proposition. Let $\mathbf{L} = \frac{1}{\sqrt{2}}\begin{pmatrix}\mathbf{I} & \mathbf{iI}\\ \mathbf{iI} & \mathbf{I}\end{pmatrix}$, $\mathbf{M}_0 = \begin{pmatrix}\mathbf{A}_0 & \mathbf{B}_0\\ \mathbf{C}_0 & \mathbf{D}_0\end{pmatrix} = \mathbf{L}^{-1}\mathbf{M}\mathbf{L}$, and let $\mathbf{Z}_0 \in \mathfrak{B}$, then

$$\begin{aligned}\mathbf{M}_0\begin{pmatrix}\mathbf{Z}_0\\ \mathbf{I}\end{pmatrix} &= \mathbf{L}^{-1}\mathbf{M}\mathbf{L}\begin{pmatrix}\mathbf{Z}_0\\ \mathbf{I}\end{pmatrix}\\ &= \frac{1}{\sqrt{2}}\mathbf{L}^{-1}\mathbf{M}\begin{pmatrix}\mathbf{Z}_0+\mathbf{iI}\\ \mathbf{iZ}_0+\mathbf{I}\end{pmatrix}\\ &= \frac{1}{\sqrt{2}}\mathbf{L}^{-1}\mathbf{M}\begin{pmatrix}(\mathbf{Z}_0+\mathbf{iI})(\mathbf{iZ}_0+\mathbf{I})^{-1}\\ \mathbf{I}\end{pmatrix}(\mathbf{iZ}_0+\mathbf{I}).\end{aligned}$$

Let $\mathbf{Z} = (\mathbf{Z}_0+\mathbf{iI})(\mathbf{iZ}_0+\mathbf{I})^{-1}$, then $\mathbf{Z} \in \mathfrak{U}$ and we have

$$\begin{aligned}\mathbf{M}_0\begin{pmatrix}\mathbf{Z}_0\\ \mathbf{I}\end{pmatrix} &= \frac{1}{\sqrt{2}}\mathbf{L}^{-1}\mathbf{M}\begin{pmatrix}\mathbf{Z}\\ \mathbf{I}\end{pmatrix}(\mathbf{iZ}_0+\mathbf{I})\\ &= \frac{1}{\sqrt{2}}\mathbf{L}^{-1}\begin{pmatrix}\mathbf{AZ}+\mathbf{B}\\ \mathbf{CZ}+\mathbf{D}\end{pmatrix}(\mathbf{iZ}_0+\mathbf{I})\\ &= \frac{1}{\sqrt{2}}\mathbf{L}^{-1}\begin{pmatrix}(\mathbf{AZ}+\mathbf{B})(\mathbf{CZ}+\mathbf{D})^{-1}\\ \mathbf{I}\end{pmatrix}(\mathbf{CZ}+\mathbf{D})(\mathbf{iZ}_0+\mathbf{I}).\end{aligned}$$

Let $\mathbf{Z}' = (\mathbf{AZ}+\mathbf{B})(\mathbf{CZ}+\mathbf{D})^{-1}$, then $\mathbf{Z}' \in \mathfrak{U}$ and we have

$$\begin{aligned}\mathbf{M}_0\begin{pmatrix}\mathbf{Z}_0\\ \mathbf{I}\end{pmatrix} &= \frac{1}{\sqrt{2}}\mathbf{L}^{-1}\begin{pmatrix}\mathbf{Z}'\\ \mathbf{I}\end{pmatrix}(\mathbf{CZ}+\mathbf{D})(\mathbf{iZ}_0+\mathbf{I})\\ &= \frac{1}{2}\begin{pmatrix}\mathbf{Z}'-\mathbf{iI}\\ \mathbf{I}-\mathbf{iZ}'\end{pmatrix}(\mathbf{CZ}+\mathbf{D})(\mathbf{iZ}_0+\mathbf{I})\\ &= \frac{1}{2}\begin{pmatrix}(\mathbf{Z}'-\mathbf{iI})(\mathbf{I}-\mathbf{iZ}')^{-1}\\ \mathbf{I}\end{pmatrix}(\mathbf{I}-\mathbf{iZ}')(\mathbf{CZ}+\mathbf{D})(\mathbf{iZ}_0+\mathbf{I}).\end{aligned}$$

Let $\mathbf{Z}'_0 = (\mathbf{Z}'-\mathbf{iI})(\mathbf{I}-\mathbf{iZ}')^{-1}$, then $\mathbf{Z}'_0 \in \mathfrak{B}$ and we have

$$\begin{aligned}\mathbf{M}_0\begin{pmatrix}\mathbf{Z}_0\\ \mathbf{I}\end{pmatrix} &= \frac{1}{2}\begin{pmatrix}\mathbf{Z}'_0\\ \mathbf{I}\end{pmatrix}(\mathbf{I}-\mathbf{iZ}')(\mathbf{CZ}+\mathbf{D})(\mathbf{iZ}_0+\mathbf{I})\\ &= \frac{1}{2}\begin{pmatrix}\mathbf{Z}'_0\\ \mathbf{I}\end{pmatrix}(\mathbf{CZ}+\mathbf{D}-\mathbf{iZ}'(\mathbf{CZ}+\mathbf{D}))(\mathbf{iZ}_0+\mathbf{I})\\ &= \frac{1}{2}\begin{pmatrix}\mathbf{Z}'_0\\ \mathbf{I}\end{pmatrix}(\mathbf{CZ}+\mathbf{D}-\mathbf{i}(\mathbf{AZ}+\mathbf{B}))(\mathbf{iZ}_0+\mathbf{I})\\ &= \frac{1}{2}\begin{pmatrix}\mathbf{Z}'_0\\ \mathbf{I}\end{pmatrix}((\mathbf{C}-\mathbf{iA})\mathbf{Z}+\mathbf{D}-\mathbf{iB})(\mathbf{iZ}_0+\mathbf{I})\\ &= \frac{1}{2}\begin{pmatrix}\mathbf{Z}'_0\\ \mathbf{I}\end{pmatrix}((\mathbf{C}-\mathbf{iA})(\mathbf{Z}_0+\mathbf{i})+(\mathbf{D}-\mathbf{iB})(\mathbf{iZ}_0+\mathbf{I}))\\ &= \frac{1}{2}\begin{pmatrix}\mathbf{Z}'_0\\ \mathbf{I}\end{pmatrix}((\mathbf{C}-\mathbf{iA}+\mathbf{B}+\mathbf{iD})\mathbf{Z}_0+\mathbf{A}+\mathbf{iC}+\mathbf{D}-\mathbf{iB})\\ &= \begin{pmatrix}\mathbf{Z}'_0\\ \mathbf{I}\end{pmatrix}(\mathbf{C}_0\mathbf{Z}_0+\mathbf{D}_0).\end{aligned}$$

From the above chain of equations, we obtain $\mathbf{C}_0\mathbf{Z}_0+\mathbf{D}_0 = \frac{1}{2}(\mathbf{I}-\mathbf{iZ}')(\mathbf{CZ}+\mathbf{D})(\mathbf{iZ}_0+\mathbf{I})$ and $\mathbf{A}_0\mathbf{Z}_0+\mathbf{B}_0 = \mathbf{Z}'_0(\mathbf{C}_0\mathbf{Z}_0+\mathbf{D}_0)$. Thus $\mathbf{C}_0\mathbf{Z}_0+\mathbf{D}_0$ is invertible and $\mathbf{Z}'_0 = (\mathbf{A}_0\mathbf{Z}_0+\mathbf{B}_0)(\mathbf{C}_0\mathbf{Z}_0+\mathbf{D}_0)^{-1}$. This shows that the Möbius transformation associated with $\mathbf{M}_0$ maps any point $\mathbf{Z}_0 \in \mathfrak{B}$ to a point $\mathbf{Z}'_0 \in \mathfrak{B}$. This completes the proof.

□

E.3 Proof of Proposition 3.3

Proof. Let $\mathbf{z} = \phi_{\mathbf{X}}(\mathbf{Y})$. Then

$$\begin{aligned}\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z})\mathbf{Z} = (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}} - \\ (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{X})^{-1}(\bar{\sigma}_{\mathbb{C}}(\mathbf{Y}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{X}))(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}(\mathbf{Y} - \mathbf{X}) \\ (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y})^{-1}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}}.\end{aligned}$$

Therefore

$$\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Z})\mathbf{Z} = (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{X})^{-1}A(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y})^{-1}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}}, \quad (12)$$

where $A = (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{X})(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y}) - (\bar{\sigma}_{\mathbb{C}}(\mathbf{Y}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{X}))(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}(\mathbf{Y} - \mathbf{X})$.

Note that

$$\begin{aligned}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{X})(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y}) = (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1} - (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y} \\ - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{X}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1} \\ + \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{X}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y},\end{aligned}$$

$$\begin{aligned}(\bar{\sigma}_{\mathbb{C}}(\mathbf{Y}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{X}))(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}(\mathbf{Y} - \mathbf{X}) = \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{Y} \\ - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{X} \\ - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})(\mathbf{I} - x\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{Y} \\ + \bar{\sigma}_{\mathbb{C}}(\mathbf{X})(\mathbf{I} - x\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{X}.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbf{A} = \ & \left((\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{X}\right) \\ & - \left((\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})(\mathbf{I} - x\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\right)\mathbf{Y} \\ & - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\left(\mathbf{X}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1} - (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{X}\right) \\ & - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\left((\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1} - \mathbf{X}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{X})\right)\mathbf{Y}.\end{aligned}$$

It is easy to see that

$$\begin{aligned}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})(\mathbf{I} - x\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{X} = \mathbf{I}, \\ (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}) - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1} = \mathbf{0}, \\ \mathbf{X}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1} - (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{X} = \mathbf{0}, \\ (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1} - \mathbf{X}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}) = \mathbf{I}.\end{aligned}$$

Therefore

$$\mathbf{A} = \mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{Y}. \quad (13)$$

Combining Eqs. (12) and (13) leads to the conclusion of the proposition.

□

E.4 Proof of Corollary 3.4

Proof. Let $\mathbf{L} = \frac{1}{\sqrt{2}}\begin{pmatrix}\mathbf{I} & \mathbf{iI} \\ \mathbf{iI} & \mathbf{I}\end{pmatrix}$, $\mathbf{M} \in \mathrm{Sp}_2(A_{\mathbb{C}}, \bar{\sigma}_{\mathbb{C}})$, and $\mathbf{M}_0 = \mathbf{L}^{-1}\mathbf{M}\mathbf{L} = \begin{pmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{pmatrix}$, then

$$\bar{\sigma}_{\mathbb{C}}(\mathbf{M})^{\mathrm{T}}\Omega\mathbf{M} = \bar{\sigma}_{\mathbb{C}}(\mathbf{L}^{-1})^{\mathrm{T}}\bar{\sigma}_{\mathbb{C}}(\mathbf{M}_0)^{\mathrm{T}}\bar{\sigma}_{\mathbb{C}}(\mathbf{L})^{\mathrm{T}}\Omega\mathbf{L}\mathbf{M}_0\mathbf{L}^{-1}.$$

Since $\bar{\sigma}_{\mathbb{C}}(\mathbf{L})^{\mathrm{T}}\Omega\mathbf{L} = \mathbf{i}\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & -\mathbf{I}\end{pmatrix}$ and $\bar{\sigma}_{\mathbb{C}}(\mathbf{M})^{\mathrm{T}}\Omega\mathbf{M} = \Omega$, we have

$$\bar{\sigma}_{\mathbb{C}}(\mathbf{M}_0)^{\mathrm{T}}\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & -\mathbf{I}\end{pmatrix}\mathbf{M}_0 = \begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & -\mathbf{I}\end{pmatrix}. \tag{14}$$

By Proposition 3.2, $(\mathbf{AY}+\mathbf{B})(\mathbf{CY}+\mathbf{D})^{-1} \in \mathfrak{B}$ for any $\mathbf{Y} \in \mathfrak{B}$. Now, let $\mathbf{X} \in \mathfrak{B}$ such that $\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}) \in (A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$, and let $\mathbf{A} = (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-\frac{1}{2}}$, $\mathbf{B} = -(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-\frac{1}{2}}\mathbf{X}$, $\mathbf{C} = -(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}}\bar{\sigma}_{\mathbb{C}}(\mathbf{X})$, $\mathbf{D} = (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}}$. Then

$$\begin{aligned}
\bar{\sigma}_{\mathbb{C}}(\mathbf{A})\mathbf{A} &= (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1},\\
\bar{\sigma}_{\mathbb{C}}(\mathbf{B})\mathbf{B} &= \bar{\sigma}_{\mathbb{C}}(\mathbf{X})(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{X},\\
\bar{\sigma}_{\mathbb{C}}(\mathbf{C})\mathbf{C} &= \mathbf{X}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}),\\
\bar{\sigma}_{\mathbb{C}}(\mathbf{D})\mathbf{D} &= (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1},\\
\bar{\sigma}_{\mathbb{C}}(\mathbf{A})\mathbf{B} &= -(\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-1}\mathbf{X},\\
\bar{\sigma}_{\mathbb{C}}(\mathbf{C})\mathbf{D} &= -\mathbf{X}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-1}.
\end{aligned}$$

It is easy to see that $\bar{\sigma}_{\mathbb{C}}(\mathbf{A})\mathbf{A} - \bar{\sigma}_{\mathbb{C}}(\mathbf{C})\mathbf{C} = \mathbf{I}$, $\bar{\sigma}_{\mathbb{C}}(\mathbf{B})\mathbf{B} - \bar{\sigma}_{\mathbb{C}}(\mathbf{D})\mathbf{D} = -\mathbf{I}$, and $\bar{\sigma}_{\mathbb{C}}(\mathbf{A})\mathbf{B} - \bar{\sigma}_{\mathbb{C}}(\mathbf{C})\mathbf{D} = \mathbf{0}$, which shows that $\mathbf{M}_0$ satisfies Eq. (14). Note that the action of $\mathbf{M}_0$ on $\mathbf{Y} \in \mathfrak{B}$ results in the map $\phi_{\mathbf{X}}(\mathbf{Y})$ given by

$$\phi_{\mathbf{X}}(\mathbf{Y}) = (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-\frac{1}{2}}(\mathbf{Y} - \mathbf{X})(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y})^{-1}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}}.$$

By Proposition 3.3, we can remark that

$$\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\phi_{\mathbf{X}}(\mathbf{Y}))\phi_{\mathbf{X}}(\mathbf{Y}) = \bar{\sigma}_{\mathbb{C}}(\mathbf{U})\mathbf{U},$$

where $\mathbf{U} = (\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{Y})\mathbf{Y})^{\frac{1}{2}}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y})^{-1}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}}$. This completes the proof of the corollary.

□

E.5 Proof of Proposition 3.5

Proof. Given

$$\mathbf{Z} = (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-\frac{1}{2}}(\mathbf{Y} - \mathbf{X})(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y})^{-1}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{\frac{1}{2}}.$$

Assume that $\mathbf{I} + \mathbf{Z}f(\mathbf{X}) \in A_{\mathbb{C}}^{\times}$, we shall prove that

$$\mathbf{Y} = (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}(\mathbf{I} + \mathbf{Z}f(\mathbf{X}))^{-1}(\mathbf{Z} + g(\mathbf{X}))(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}}.$$

Setting $\tilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}\mathbf{Z}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}}$, we have

$$\tilde{\mathbf{Z}} = (\mathbf{Y} - \mathbf{X})(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{Y})^{-1}.$$

Therefore

$$(\tilde{\mathbf{Z}} + \mathbf{X}) = (\mathbf{I} + \tilde{\mathbf{Z}}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))\mathbf{Y}. \tag{15}$$

Note that

$$\begin{aligned}
\tilde{\mathbf{Z}} + \mathbf{X} &= (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}\mathbf{Z}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}} + \mathbf{X}\\
&= (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}\mathbf{Z}(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}} + (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}g(\mathbf{X})(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}} \quad (16)\\
&= (\mathbf{I} - \mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}(\mathbf{Z} + g(\mathbf{X}))(\mathbf{I} - \bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}}.
\end{aligned}$$

Also, we have

$$\begin{aligned}\mathbf{I}+\tilde{\mathbf{Z}}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}) &= \mathbf{I}+(\mathbf{I}-\mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}\mathbf{Z}(\mathbf{I}-\bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}) \\ &= \mathbf{I}+(\mathbf{I}-\mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}\mathbf{Z}f(\mathbf{X})(\mathbf{I}-\mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-\frac{1}{2}} \\ &= (\mathbf{I}-\mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}(\mathbf{I}+\mathbf{Z}f(\mathbf{X}))(\mathbf{I}-\mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{-\frac{1}{2}}.\end{aligned} \tag{17}$$

Combining Eqs. (15), (16), and (17), we obtain

$$\mathbf{Y}=(\mathbf{I}-\mathbf{X}\bar{\sigma}_{\mathbb{C}}(\mathbf{X}))^{\frac{1}{2}}(\mathbf{I}+\mathbf{Z}f(\mathbf{X}))^{-1}(\mathbf{Z}+g(\mathbf{X}))(\mathbf{I}-\bar{\sigma}_{\mathbb{C}}(\mathbf{X})\mathbf{X})^{-\frac{1}{2}}.$$

□

E.6 Proof of Proposition 3.6

Proof. First, note that

$$\rho(\mathbf{Z})=\begin{pmatrix}\mathbf{I} & \mathbf{X}_0\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{Y}_0 & \mathbf{0}\\ \mathbf{0} & \mathbf{Y}_0^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \bar{\sigma}_{\mathbb{C}}(\mathbf{X}_0) & \mathbf{I}\end{pmatrix}. \tag{18}$$

If $\mathbf{Z}\in\mathfrak{U}$, then $\mathbf{Y}_0\in(A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$ and there exists $\mathbf{P}\in A_{\mathbb{C}}^{\times}$ such that $\mathbf{Y}_0=\mathbf{P}\bar{\sigma}_{\mathbb{C}}(\mathbf{P})$. Thus

$$\begin{aligned}\rho(\mathbf{Z}) &= \begin{pmatrix}\mathbf{I} & \mathbf{X}_0\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{P}\bar{\sigma}_{\mathbb{C}}(\mathbf{P}) & \mathbf{0}\\ \mathbf{0} & \bar{\sigma}_{\mathbb{C}}(\mathbf{P})^{-1}\mathbf{P}^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \bar{\sigma}_{\mathbb{C}}(\mathbf{X}_0) & \mathbf{I}\end{pmatrix}\\ &= \mathbf{Q}\bar{\sigma}_{\mathbb{C}}(\mathbf{Q}),\end{aligned}$$

where $\mathbf{Q}=\begin{pmatrix}\mathbf{I} & \mathbf{X}_0\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{P} & \mathbf{0}\\ \mathbf{0} & \bar{\sigma}_{\mathbb{C}}(\mathbf{P})^{-1}\end{pmatrix}\in A_{\mathbb{C}}^{\times}$. Therefore, $\rho(\mathbf{Z})\in(A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$.

If $\rho(\mathbf{Z})\in(A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$, then there exists $\mathbf{P}\in A_{\mathbb{C}}^{\times}$ such that $\rho(\mathbf{Z})=\mathbf{P}\bar{\sigma}_{\mathbb{C}}(\mathbf{P})$. By Eq. (18),

$$\begin{aligned}\begin{pmatrix}\mathbf{Y}_0 & \mathbf{0}\\ \mathbf{0} & \mathbf{Y}_0^{-1}\end{pmatrix} &= \begin{pmatrix}\mathbf{I} & \mathbf{X}_0\\ \mathbf{0} & \mathbf{I}\end{pmatrix}^{-1}\mathbf{P}\bar{\sigma}_{\mathbb{C}}(\mathbf{P})\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \bar{\sigma}_{\mathbb{C}}(\mathbf{X}_0) & \mathbf{I}\end{pmatrix}^{-1}\\ &= \mathbf{Q}\bar{\sigma}_{\mathbb{C}}(\mathbf{Q}),\end{aligned}$$

where $\mathbf{Q}=\begin{pmatrix}\mathbf{I} & \mathbf{X}_0\\ \mathbf{0} & \mathbf{I}\end{pmatrix}^{-1}\mathbf{P}\in A_{\mathbb{C}}^{\times}$. Let $\mathbf{Q}=\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}$. Then $\mathbf{Y}_0=\mathbf{A}\bar{\sigma}_{\mathbb{C}}(\mathbf{A})+\mathbf{B}\bar{\sigma}_{\mathbb{C}}(\mathbf{B})$. Since $\mathbf{Y}_0\in A_{\mathbb{C}}^{\times}$, we deduce that $\mathbf{Y}_0\in(A_{\mathbb{C}}^{\bar{\sigma}_{\mathbb{C}}})_+$, which concludes the proposition.

□

E.7 Proof of Corollary 4.1

Proof. The corollary can be proved by noting that the condition $\mathbf{I}-\mathbf{X}^H\mathbf{X}\in\mathbb{H}_n^+$ implies that $\mathbf{I}-\mathbf{X}\mathbf{X}^H\in\mathbb{H}_n^+$.

□

E.8 Proof of Corollary 4.2

Proof. It is easy to see that $\mathbf{X}^H(\mathbf{I}-\mathbf{X}\mathbf{X}^H)^{\frac{1}{2}}=(\mathbf{I}-\mathbf{X}^H\mathbf{X})^{\frac{1}{2}}\mathbf{X}^H$ and $\mathbf{X}(\mathbf{I}-\mathbf{X}^H\mathbf{X})^{\frac{1}{2}}=(\mathbf{I}-\mathbf{X}\mathbf{X}^H)^{\frac{1}{2}}\mathbf{X}$. Therefore, $f(\mathbf{X})=\mathbf{X}^H$, $g(\mathbf{X})=\mathbf{X}$, and we obtain the desired formula.

□

E.9 Proof of Proposition 5.1

Proof. Let $\beta=\alpha_{\mathbf{X},\mathbf{Y}}(t)$, $\mathbf{Z}=\phi_{\mathbf{X}}(\mathbf{Y})$, and let $\mathbf{Z}=\mathbf{U}\mathbf{D}\mathbf{V}$ be the singular value decomposition of $\mathbf{Z}$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, and $\mathbf{D}$ is a real diagonal matrix. Then

$$(\mathbf{I}-\beta^2\mathbf{Z}\mathbf{Z}^H)^{-\frac{1}{2}}=\mathbf{U}(\mathbf{I}-\beta^2\mathbf{D}^2)^{-\frac{1}{2}}\mathbf{U}^{-1},$$

$$(\mathbf{I} - \beta \mathbf{Z}^H \mathbf{Z})^{-1} = \mathbf{V}^{-1}(\mathbf{I} - \beta \mathbf{D}^2)^{-1}\mathbf{V},$$
$$(\mathbf{I} - \beta^2 \mathbf{Z}^H \mathbf{Z})^{\frac{1}{2}} = \mathbf{V}^{-1}(\mathbf{I} - \beta^2 \mathbf{D}^2)^{\frac{1}{2}}\mathbf{V}.$$

Thus

$$\begin{aligned}\phi_{\beta\mathbf{Z}}(\mathbf{Z}) &= \mathbf{U}(\mathbf{I} - \beta^2\mathbf{D}^2)^{-\frac{1}{2}}\mathbf{U}^{-1}(1-\beta)\mathbf{U}\mathbf{D}\mathbf{V}\mathbf{V}^{-1}(\mathbf{I} - \beta\mathbf{D}^2)^{-1}\mathbf{V}\mathbf{V}^{-1}(\mathbf{I} - \beta^2\mathbf{D}^2)^{\frac{1}{2}}\mathbf{V} \\ &= (1-\beta)\mathbf{U}\mathbf{D}(\mathbf{I} - \beta\mathbf{D}^2)^{-1}\mathbf{V}.\end{aligned}$$

According to the definition of the spectral norm $\|\cdot\|_2$,

$$\|\phi_{\beta\mathbf{Z}}(\mathbf{Z})\|_2 = \frac{(1-\beta)\|\mathbf{Z}\|_2}{1 - \beta\|\mathbf{Z}\|_2^2}.$$

Therefore

$$\begin{aligned}\frac{1 + \|\phi_{\beta\mathbf{Z}}(\mathbf{Z})\|_2}{1 - \|\phi_{\beta\mathbf{Z}}(\mathbf{Z})\|_2} &= \frac{1 + \|\mathbf{Z}\|_2 - \beta\|\mathbf{Z}\|_2 - \beta\|\mathbf{Z}\|_2^2}{1 - \|\mathbf{Z}\|_2 + \beta\|\mathbf{Z}\|_2 - \beta\|\mathbf{Z}\|_2^2} \\ &= \frac{1 - \beta\|\mathbf{Z}\|_2}{1 + \beta\|\mathbf{Z}\|_2}\frac{1 + \|\mathbf{Z}\|_2}{1 - \|\mathbf{Z}\|_2},\end{aligned}$$

which results in

$$\log\left(\frac{1 + \beta\|\mathbf{Z}\|_2}{1 - \beta\|\mathbf{Z}\|_2}\right) + \log\left(\frac{1 + \|\phi_{\beta\mathbf{Z}}(\mathbf{Z})\|_2}{1 - \|\phi_{\beta\mathbf{Z}}(\mathbf{Z})\|_2}\right) = \log\left(\frac{1 + \|\mathbf{Z}\|_2}{1 - \|\mathbf{Z}\|_2}\right),$$

or, equivalently,

$$d^K(\mathbf{0}_n, \beta\mathbf{Z}) + d^K(\beta\mathbf{Z}, \mathbf{Z}) = d^K(\mathbf{0}_n, \mathbf{Z}).$$

The last equality shows that $\alpha_{\mathbf{X},\mathbf{Y}}(t)\phi_{\mathbf{X}}(\mathbf{Y})$ must lie on a minimizing geodesic between $\mathbf{0}_n$ and $\phi_{\mathbf{X}}(\mathbf{Y})$. The conclusion of the proposition then follows from Eqs. (5) and (6).

□