SPD-DDPM: Denoising Diffusion Probabilistic Models in the Symmetric Positive Definite Space

Yunchen Li\textsuperscript{1*}, Zhou Yu\textsuperscript{1,2*}, Gaoqi He\textsuperscript{1}, Yunhang Shen\textsuperscript{3}, Ke Li\textsuperscript{3}, Xing Sun\textsuperscript{3}, Shaohui Lin\textsuperscript{1,2†}

\textsuperscript{1}East China Normal University
\textsuperscript{2}Key Laboratory of Advanced Theory and Application in Statistics and Data Science, Ministry of Education, China
\textsuperscript{3}YouTu Lab, Tencent, Shanghai, China

Corresponding author
These authors contributed equally.
\{shenyunhang01, tristanli.sh, winfred.sun, shaohuilin007\}@gmail.com

Abstract
Symmetric positive definite (SPD) matrices have shown important value and applications in statistics and machine learning, such as FMRI analysis and traffic prediction. Previous works on SPD matrices mostly focus on discriminative models, where predictions are made directly on \(E(X|y)\), by introducing Gaussian distribution in the SPD space to estimate \(E(X|y)\). Moreover, our model can estimate \(p(X)\) unconditionally and flexibly without giving \(y\). On the one hand, the model conditionally learns \(p(X|y)\) and utilizes the mean of samples to obtain \(E(X|y)\) as a prediction. On the other hand, the model unconditionally learns the probability distribution of the data \(p(X)\) and generates samples that conform to this distribution. Furthermore, we propose a new SPD net which is much deeper than the previous networks and allows for the inclusion of conditional factors. Experiment results on toy data and real taxi data demonstrate that our models effectively fit the data distribution both unconditionally and conditionally.

Introduction
Symmetric positive definite (SPD) matrices hold significant importance within the domains of machine learning and multivariate statistical analysis. They play a pervasive role across a diverse range of applications, including FMRI analysis (Petersen and Müller 2019), hand gesture recognition (Nguyen et al. 2019), facial recognition (Otterdou et al. 2019), traffic prediction (Tucker, Wu, and Müller 2023), brain-computer interfaces (Barachant et al. 2013), video classification (Wang et al. 2023a; Huang and Van Gool 2017; Wang et al. 2022), interaction recognition (Nguyen 2021), and many more.

There are two main tasks for SPD matrices: classification and SPD matrix estimation. The first one involves using SPD matrices as inputs to predict the classes of matrices. They often involve transforming SPD matrices into Euclidean space vectors via a Log mapping and then employ SPDNet accompanied by convolution with activation (Higham 2008) and batch normalization (Brooks et al. 2019) for classification. Recently, several works have been proposed to modify SPDNet via multi-scale features, UNet structure, and local convolution (Chen et al. 2023; Wang et al. 2022; Zhang et al. 2020). The second one aims to estimate the SPD matrix as the predicted variable \(X\). Previous methods often use traditional statistical modeling frameworks as discriminative models by estimating \(E(X|y)\) given the conditional variables \(y\). For example, they define an additive model in a reproducing kernel Hilbert space using kernel functions (Lin, Müller, and Park 2023) or define a regression model in a metric space upon Fréchet mean (Petersen and Müller 2019; Qiu, Yu, and Zhu 2022). However, these models often face challenges when dealing with high-dimensional SPD matrices or predictor \(y\). In addition, these methods primarily focus on predicting \(E(X|y)\) and currently lack approaches for estimating the probability density in the SPD space. This motivates the construction of generative models for the SPD space. Such models not only benefit from predicting \(p(X)\) but also utilize maximum likelihood estimation to make predictions for \(X\) given conditional variables \(y\).

Denoising diffusion probabilistic model (DDPM) (Ho, Jain, and Abbeel 2020) is one of the most popular and effective generative models. Several researchers have extended DDPM to Riemannian manifolds, including SO(3) (Leach et al. 2022), SE(3) (Yim et al. 2023), spheres (Huang et al. 2022), and more (De Bortoli et al. 2022). These extensions have proven to be of practical value in fields such as protein generation (Corso et al. 2022), molecular docking (Corso et al. 2022), and others. However, few works have discussed and studied the SPD space. Meanwhile, the classical DDPM has poor performance when estimating probability distribution in the SPD space. When real data resides on a low-dimensional manifold, estimating in a high-dimensional Euclidean space is inefficient, and the distribution on the manifold may not exhibit favorable characteristics in the Euclidean space where it is embedded.

To address the SPD space distribution estimation prob-
lem, we propose a novel denoising diffusion probabilistic model in the SPD space, termed SPD-DDPM. The key components of SPD-DDPM lie in the equation of the backward process and the network to generate SPD matrices. Firstly, by introducing Gaussian distribution, addition and multiplication operations in the SPD space, we extend the DDPM in the Euclidean space to the SPD space, including the forward and backward process. Secondly, we introduce a novel SPD net that takes an SPD matrix as input using double convolutions and allowing for the incorporation of conditions into the network.

Our contributions are summarized as follows:

- We propose a novel SPD-DDPM to accurately estimate the SPD matrix $X$, which provides two versions: conditional and unconditional generation.
- We introduce the Gaussian distribution in the SPD space, the addition and multiplication operations in the SPD space to extend the DDPM theory.
- A new SPD U-Net is proposed to effectively incorporate the conditions during the generative process.

Experiment results on toy data and real taxi data demonstrate that our models effectively fit the data distribution both unconditionally and conditionally and provide accurate predictions.

**Preliminary**

**Denoising Diffusion Probabilistic Model (DDPM)**

Assume the source data $X$ is a distribution $X \sim q(X)$. DDPM defines a forward process of $T$ steps that gradually adds noise to transform the source data into a standard Gaussian distribution by:

$$q(X_t | X_{t-1}) = N(\sqrt{1 - \beta_t} X_{t-1}, \beta_t I),$$  

where $X_t$ is the forward output at the $t$-th step, and $\beta_t$ is a hyperparameter. Eq. 1 can be reparameterized as:

$$X_t = \sqrt{1 - \beta_t} X_{t-1} + \sqrt{\beta_t} \epsilon_t, \ \epsilon_t \sim N(0, I).$$  

After that, we build the distribution of $q(X_t | X_0)$ as:

$$q(X_t | X_0) = N(\sqrt{\alpha_t} X_0, (1 - \sqrt{\alpha_t}) I),$$  

$$\alpha_t = 1 - \beta_t, \ \bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i.$$

Note that the forward process does not require training. After $T$ forward steps, the source data $X$, a.k.a. $X_0$, is transformed into a normal distribution.

In the backward process, DDPM attempts to solve $q(X_{t-1} | X_t, X_0)$ based on Bayes theorem:

$$q(X_{t-1} | X_t, X_0) = \frac{q(X_t | X_{t-1}) q(X_{t-1} | X_0)}{q(X_t | X_0)}$$  

$$= N(\mu_t(X_t, X_0), \sigma_t^2 I),$$  

where $\epsilon \sim N(0, I)$ represents the noise added to $X_t$, but we do not know its exact value when given $X_t$. Thus, a neural network $\epsilon_\theta$ is used to approximate it. The loss function can be formulated as:

$$\|\epsilon - \epsilon_\theta(X_t, t)\|^2$$

According to Eq. 4, the reverse process is constructed to obtain $p(\bar{X}_{t-1} | \bar{X}_t)$:

$$p(X_{t-1} | X_t) = N(\frac{1}{\sqrt{\alpha_t}} X_t - \frac{1 - \alpha_t}{\sqrt{\alpha_t}(1 - \alpha_t)} \epsilon_\theta, \Sigma_t),$$  

where $\Sigma_t = \sigma_t^2 I$.

During inference, the data $X_T$ is sampled from the standard Gaussian distribution and then employ Eq. 6 to gradually remove noise to generate the desired output of $\bar{X}$.

**Symmetric Positive Definite (SPD) Space**

Various real data (e.g., FRMI) are satisfied with the SPD condition to form SPD matrices, which is significantly different from the data in the Euclidean space. In the SPD space (denoted by $Sym^+$), symmetry and positive definiteness are held:

$$Symmetry: \forall X \in Sym^+, X^T = X,$$

$$Positive: \forall X \in Sym^+, \forall a \in R^n, a^T X a > 0.$$

Following (Said et al. 2017), Gaussian distribution in the SPD space is defined as:

$$p(X | \bar{X}, \sigma^2) = \frac{1}{\zeta(\sigma)} \exp[-\frac{d(X, \bar{X})^2}{2\sigma^2}],$$

where $\zeta(\sigma)$ is the regularization coefficient. The distance $d$ in Eq. 8 is affine-invariant metric (Moakher 2005), which is defined as:

$$d(X_1, X_2)^2 = tr[\log((X_1^{-0.5} X_2 X_1^{-0.5}))^2].$$

Based on the above definition of affine-invariant metric, the exponent and logarithm mappings are formulated by following the work (Higham 2008):

$$\text{Exp}_{X_1}(X_2) = X_1^{\frac{1}{2}} \exp(X_1^{-\frac{1}{2}} X_2 X_1^{-\frac{1}{2}}) X_1^{\frac{1}{2}},$$

$$\text{Log}_{X_1}(X_2) = X_1^{\frac{1}{2}} \log(X_1^{-\frac{1}{2}} X_2 X_1^{-\frac{1}{2}}) X_1^{\frac{1}{2}}.$$

In the SPD space, we can employ the exponent and logarithm mappings in Eq. 10 to further define the addition and multiplication operations in SPD space for computation:

In the SPD space, the base matrix we select is identity matrix $I$, and the addition operation $\oplus$ and multiplication operation $\odot$ can be defined as:

**Definition 1**

$$X_1 \oplus X_2 = \text{Exp}_I(\text{Log}_I(X_1) + \text{Log}_I(X_2)).$$

$$r \odot X = \text{Exp}_I(r \cdot \text{Log}_I(X)).$$

Since the base matrix $I$ used in Eq. 10, the exponent and logarithm mappings are equal to $\exp$ and $\log$ function. Thus, addition and multiplication operations in Definition 1 are reformulated as:

$$X_1 \oplus X_2 = \exp(\log(X_1) + \log(X_2))$$

$$= \exp(\frac{1}{2} \log(X_1) + \frac{1}{2} \log(X_2) + \frac{1}{2} \log(X_1))$$

$$= \exp(\log(X_1^\frac{1}{2}) + \log(X_2) + \log(X_1^\frac{1}{2}))$$

$$= \exp(\log(X_1^\frac{1}{2} X_2 X_1^\frac{1}{2})) = X_1^\frac{1}{2} X_2 X_1^\frac{1}{2},$$

$$r \odot X = \exp(r \cdot \log(X)) = \exp(\log(X^r)) = X^r.$$
In the work (Terras 2012), SPD space with affine-invariant metric is a homogeneous space under the action of the linear group $GL(m)$, where the group action is defined as:

$$(X, A) \rightarrow X \cdot A \overset{def}{=} A \cdot X A.$$  

(14)

Therefore, the affine-invariant metric remains invariant under group actions:

$$d(X_1, X_2)^2 = d(X_1 \cdot A, X_2 \cdot A)^2.$$  

(15)

Based on Eqs. 15, 8, and 9, we have the following proposition:

**Proposition 1** $X \sim G(\bar{X}, \sigma^2) \Rightarrow X \cdot A \sim G(\bar{X} \cdot A, \sigma^2)$. 

Proof:

For any function $f : P_m \rightarrow R$ ($P_m$ is SPD space), and any $A \in GL(m)$, according to Eq. 14, we have:

$$\int_{P_m} f(X) dv(X) = \int_{P_m} f(X \cdot A) dv(X).$$

where $dv(X)$ is the Riemannian volume element defined as:

$$dv(X) = det(X)^{-\frac{m+1}{2}} \prod_{i < j} dX_{ij}.$$ 

Intuitively, $dv(X)$ is the Riemannian volume element of $ds^2(X)$, and $ds^2(X)$ is invariant under congruence transformations and inversion:

$$ds^2(X) = tr[X^{-1} dX]^2.$$ 

Let $X$ be a random variable in $P_m$, let $\phi : P_m \rightarrow R$ be a test function. If $X \sim G(\bar{X}, \sigma^2)$ and $Z = X \cdot A$, then the expectation of $\phi(Z)$ is given by:

$$\int_{P_m} \phi(X \cdot A)p(X | \bar{X}, \sigma^2)dv(X) = \int_{P_m} \phi(Z)p(Z | A^{-1} \bar{X}, \sigma^2)dv(Z) = \int_{P_m} \phi(Z)p(Z | \bar{X}, \sigma^2)dv(Z).$$

### Method

**Symmetric Positive Definite Denoising Diffusion Probabilistic Model (SDP-DDPM)**

Different from DDPM, which is computed in Euclidean space, we propose SDP-DDPM based on the operational rules (i.e., Eq. 12 and Eq. 13) and Proposition 1 in the SPD space.

In the forward process, following DDPM, we can formulate $q(X_t | X_{t-1})$ in the SPD space as:

$$q(X_t | X_{t-1}) \sim G(\alpha_t \odot X_{t-1}, \beta_t^2),$$

(16)

where $\alpha_t^2 + \beta_t^2 = 1$, and $G(\cdot , \cdot)$ is gaussian distribution in the SPD space defined in Eq. 8. By using Eq. 12, Eq. 13) and Proposition 1, we have:

$$X_t = \alpha_t \odot X_{t-1} \odot \beta_t \odot \epsilon_t, \quad \epsilon_t \sim G(I, 1)$$

(17)

$$X_t \sim G(\alpha_t \odot X_{t-1}, \beta_t^2).$$

By using Eq. 17, we can obtain the relationship between $X_t$ and $X_0$ as below:

$$X_t = \alpha_t \odot X_{t-1} \odot \beta_t \odot \epsilon_t$$

$$= \alpha_t \odot (\alpha_{t-1} \odot X_{t-2} \odot \beta_{t-1} \odot \epsilon_{t-1}) \odot \beta_t \odot \epsilon_t$$

$$= (\alpha_t \odot \alpha_{t-1} \odot X_{t-2} \odot \alpha_{t-2} \odot \beta_{t-2} \odot \epsilon_{t-2} \odot \beta_{t-2} \odot \epsilon_{t-2})$$

(18)

$$= \alpha_t \odot X_0 \odot \sqrt{1 - \alpha_t^2} \odot \epsilon, \quad \epsilon \sim G(I, 1)$$

$$\sim G(\alpha_t \odot X_0, 1 - \alpha_t^2), \quad \alpha_t = \prod_{i=0}^{t} \alpha_i.$$ 

By defining $\beta_t$ to be $\sqrt{1 - \alpha_t^2}$, $X_t$ can be further formulated as:

$$X_t = \alpha_t \odot X_0 \odot \beta_t \odot \epsilon.$$  

(19)

Based on Eq. 19, we only set $\alpha_t$ to approach 0, $X_t$ will be transformed to the standard Gaussian distribution in the SPD space. Empirically, $\alpha_t$ is set to $\sqrt{1 - \frac{0.08t}{T}}$ in our experiments, such that $\alpha_t \rightarrow 0$ when $t \rightarrow \infty$. Therefore, we can randomly sample the noise from the standard Gaussian distribution in the SPD space, and use the following backward process to generate the desired output.

In the backward process, we also compute $q(X_{t-1} | X_t, X_0)$ using the Bayesian technique as:

$$q(X_{t-1} | X_t, X_0) = q(X_{t-1} | X_t) \frac{q(X_t | X_0)}{q(X_t | X_0)}$$

$$\propto \exp \left(- \frac{d(X_{t-1}, X_t^{\alpha_t})^2}{2\beta_t^2} - \frac{d(X_t, X_0^{\alpha_{t-1}})}{2\beta_{t-1}} \right)$$

$$\propto \exp \left[- \frac{tr(\beta_{t-1}^2 \alpha_t^2 + \beta_t^2 (\log \mu_t^{-\frac{1}{2}} X_t \mu_t^{-\frac{1}{2}}))^2}{2\beta_{t-1}^2 \beta_t^2} \right]$$

$$\sim G(\mu(X_t, \epsilon), \sigma_t^2),$$

with $\mu(x_0, \epsilon) = \frac{1}{\alpha_t} \odot X_t \odot \frac{\beta_t^2}{\alpha_t \beta_t} \odot \epsilon, \quad \sigma_t = \frac{\beta_{t-1} \beta_t}{\beta_t}.$

Note that the final computation expression of $q(X_{t-1} | X_t, X_0)$ is similar to DDPM (i.e., Eq. 4), while their difference is that our addition and scalar multiplication are defined through exponential and logarithmic mappings. In Euclidean space, the exponential and logarithmic mappings can be regarded as addition and subtraction operations.

Further, we also formulate $p(X_{t-1} | X_t)$ as:

$$p(X_{t-1} | X_t) \sim G(\mu_t(X_t, t), \Sigma_t^2),$$

(21)

where $\Sigma_t^2 = \beta_t^2 \cdot \mu_t(X_t, t)$ can be computed according to $q(X_{t-1} | X_t, X_0)$:

$$\mu_t(X_t, t) = \frac{1}{\alpha_t} \odot X_t \odot \frac{\beta_t^2}{\alpha_t \beta_t} \odot \epsilon_t.$$  

(22)

During training, we aim to minimize the KL divergence between $p(X_{t-1} | X_t)$ and $q(X_{t-1} | X_t, X_0)$. We also found that $\text{KL}(q(X_{t-1} | X_t, X_0) \parallel p(X_{t-1} | X_t))$ is bounded by
then propose the conditional process of SPD-DDPM. In the following, we will first de-

The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

13712

Algorithm 1: Unconditional Training of SPD-DDPM.

1: repeat
2: \(t \sim \text{Uniform}(1, \ldots, T)\)
3: \(X_0 \sim q(X_0)\)
4: \(\epsilon \sim G(I, 1)\)
5: \(X_t = \bar{\alpha}_t \odot X_0 \oplus \bar{\beta}_t \odot \epsilon\)
6: Take gradient descent step on \(\nabla_{\theta} \| \epsilon - \epsilon_\theta(X_t, t) \|\)
7: until convergence

Algorithm 2: Unconditional Sampling of SPD-DDPM.

1: \(X_T \sim G(I, 1)\)
2: for \(t = T, \ldots, 1\) do
3: \(z \sim G(I, 1)\)
4: \(X_{t-1} = \frac{1}{\alpha_t} \odot (X_t \oplus \frac{\beta_t}{\beta_\theta} \epsilon_\theta) \oplus \bar{\sigma}_t / \gamma \odot z\), where \(\odot\) is defined
by using \(=\) to replace \(+\) in Eq. 1.
5: end for
6: return \(X_0\)
SPD-DDPM can be formulated as:

\[ \| \epsilon - \epsilon_0(X_i, t, y) \| . \] (28)

As presented in Alg. 3, the training process is similar to the unconditional SPD-DDPM.

In conditional generation, we aim to incorporate conditional generation as a novel approach for regression modeling. In regression models, it is common to assume that \( X \) given \( y \) follows a certain distribution and use \( E(X|y) \) as the prediction of \( X \) given \( y \). In this work, we assume \( X|y \sim G(\mu_y, \sigma^2) \) with \( \mu_y \) being \( E(X|y) \). Said et al. (Said et al. 2017) have deduced that the maximum likelihood estimation of \( E(X|y) \) is the empirical Riemannian centre of samples from \( P(X|y) \), where the Riemannian centre \( \hat{X}_N \) is defined as:

\[ \hat{X}_N = \arg \min_X \frac{1}{N} \sum_{n=1}^{N} d(X, X^n)^2. \] (29)

\( \hat{X}_N \) under the affine-invariant metric in Eq. 29 can be obtained using a gradient descent algorithm (Dryden, Kolodydenko, and Zhou 2009; Pennec 1999; Pennec, Fillard, and Ayache 2006). Therefore, we train the model to fit \( p(X|y) \).

During inference, given a specific \( y \), we first generate \( N \) samples from \( p(X|y) \). We then compute \( \hat{X}_N \) by minimizing Eq. (29) as the prediction of \( E(X|y) \). The detailed inference process is presented in Alg. 4.

**Experiments**

**Experimental Setups**

**Datasets.** We use the simulation or toy data to test the performance of unconditional SPD-DDPM. We first randomly select a matrix \( A \) as the center point of the distribution, and then sample total 15,000 SPD matrices from \( G(A, \sigma^2) \) as the training data.

In the conditional generation, taxi data from the New York City Taxi and Limousine Commission is selected, which is available in ¹ Data preprocessing is following (Tucker, Wu, and Müller 2023). New York City is divided into 10 boroughs. Each element of the SPD matrix \( X_{ij} \) represents the passenger flow between the \( i \)-th borough and the \( j \)-th borough. We obtain 8,700 weighted adjacency matrices by collecting data from 2019.1 to 2019.12 and also collect the 13 predictors following (Tucker, Wu, and Müller 2023). We select 7,600 samples as the training set and 1,100 samples as the testing set.

**Evaluation metric.** In the unconditional generation, the Affine-invariant metric is widely used to test performance:

\[ d(A, B)^2 = tr[\log(A^{-0.5} B A^{-0.5})]^2. \]

In the condition generation, there is no previous method that optimizes under the Affine-invariant metric, so we select both Frobenius and Affine-invariant metric as the evaluation

¹https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page.
Table 1: Mean distance between the generated samples and the real distribution in the unconditional generation. \(a \pm b\) means the mean and std in all tables by running the model 5 times, respectively.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Distance, (\pm) std</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDPM</td>
<td>813.75 (\pm) 187.92</td>
</tr>
<tr>
<td>SPD-DDPM</td>
<td>23.17 (\pm) 3.00</td>
</tr>
</tbody>
</table>

Table 2: Comparison among Frechet Regression, DDPM and SPD-DDPM for conditional generation.

<table>
<thead>
<tr>
<th>Method</th>
<th>Frobenius (\downarrow)</th>
<th>Affine-invariant (\downarrow)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frechet Regression</td>
<td>6.6448</td>
<td>8.0099</td>
</tr>
<tr>
<td>DDPM</td>
<td>62.97</td>
<td>1203.10</td>
</tr>
<tr>
<td>SPD-DDPM</td>
<td>6.4684</td>
<td>0.6123</td>
</tr>
</tbody>
</table>

Conditional Generation

Our goal is to use predictor variables \(y\) to generate responding SPD matrices, achieving generative predictive modeling. Practically, we generate \(n = 20\) samples for each predictor \(y\). We compare our method with the discriminative method Frechet Regression (Petersen and Müller 2019) and DDPM. Our method is optimized under the Affine-invariant metric, while Frechet Regression is optimized under the Forbinus metric. The results are summarized in Tab. 2. Our method outperforms Frechet Regression in both Frobenius and Affine-invariant metrics. In the Frobenius metric, our method slightly outperforms Frechet Regression because it is optimized under the Frobenius norm. In the Affine-invariant metric, our method significantly outperforms Frechet Regression. Our conditional SPD-DDPM also achieves significantly lower Frobenius metric, compared to DDPM.

To further evaluate the effectiveness of our method, we visualize the SPD matrices in the Manhattan map Fig. 4. The left visualization represents the Frechet Regression, followed by the visualizations for our method and real data. We found that the predictions of Frechet Regression are overestimated and fail to capture the differences between each edge effectively. While our method shows a closer approximation to the real matrix.

Actually, discriminative methods require inputting all the data into the model simultaneously, rather than training in batches. When dealing with large-scale data, it will exceed memory limitations. Furthermore, their approach lacks the capability to retain intermediate model parameters, necessitating retraining the model when new samples need estimation. Our approach is based on the neural network, allowing us to circumvent the issues mentioned above.

Ablation Study

In this ablation study, we mainly evaluate the effects of hyperparameter \(\gamma\) and SPD U-Net, which are very important to SPD-DDPM.

Effect of \(\gamma\). We evaluate the effect of \(\gamma\) in unconditional generation. As shown in Tab. 3, larger \(\gamma\) generates a lower mean distance to better estimate the inputs. In addition, the std is also reduced by increasing \(\gamma\).

Effect of SPD U-Net. In this paper, we propose a high-capacity SPD U-Net. Compared to the traditional SPD net, the network proposed in this paper is deeper and incorporates a double convolution, down and up convolution, and concatenation structure. We further demonstrate the effectiveness of these modifications through ablation experiments in unconditional generation by comparing the mean distance.
Figure 4: Visualization of taxi data. On the left is the data predicted by Frechet Regression, in the middle is the data predicted by our method, and on the right is the real data. Each node represents a region, and the thickness of the line segments represents the number of taxis.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Distance ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without double convolution</td>
<td>26.17 ± 3.75</td>
</tr>
<tr>
<td>Without concatenation</td>
<td></td>
</tr>
<tr>
<td>Without down&amp;up convolution</td>
<td>42.75 ± 3.25</td>
</tr>
<tr>
<td>SPD-DDPM with SPD U-Net</td>
<td>23.17 ± 3.00</td>
</tr>
</tbody>
</table>

Table 4: Ablation Study of SPD U-Net.

Related Work

Denoising Diffusion Probabilistic Model

DDPM is a generative model proposed by Ho et al. (Ho, Jain, and Abbeel 2020). Nichol et al. (Nichol and Dhariwal 2021) and Song et al. (Song, Meng, and Ermon 2020) further made improvements to it. The diffusion model is first applied in computer vision and found numerous applications in conditional image generation (Nichol et al. 2021; Rombach et al. 2022), image restoration (Wang, Yu, and Zhang 2022; Lugmayr et al. 2022), 3D image generation (Wang et al. 2023b), and other fields. Song et al. (Song et al. 2020) introduced the score-based generation method to accelerate the generation speed of the diffusion model from the perspective of stochastic differential equations (Lu et al. 2022). Different with these methods, our SPD-DDPM is constructed on the SPD space, rather than Euclidean space. By introducing Gaussian distribution, the addition and multiplication operations in the SPD space, we effectively generate the forward and backward processes in the SPD space for better estimating the SPD matrix unconditionally and conditionally.

Symmetric Positive Definite Metric

SPD space is a nonlinear metric space. Depending on the metric, SPD forms a Riemannian manifold. Different matrices have been proposed, such as affine-invariant metric (Moakher 2005; Pennec 2006; Fletcher and Joshi 2007), log-Euclidean metric (Arsigny et al. 2007), log-Cholesky metric (Lin 2019) and scaling rotation distance (Jung, Schwartzman, and Groisser 2015). Some work studied regression with SPD matrix valued responses. Using Frechet mean (Fréchet 1948), Petersen et al. (Petersen and Müller 2019) proposed Frechet Regression under different metric. Based on this, Tucker et al. (Tucker, Wu, and Müller 2023) proposed a method for variable selection. Lin et al. (Lin, Müller, and Park 2023) proposed an adaptive model in SPD space. Qiu et al. (Qiu, Yu, and Zhu 2022) propose random forest with SPD matrix response. Different from these methods, our SPD-DDPM is the first generation model in SPD space, which can fit probability distribution in SPD space.

Conclusion

This paper proposes a novel denoising diffusion probabilistic model in the SPD space, termed SDP-DDPM. The proposed SDP-DDPM is applied for both unconditional and conditional generation. In the unconditional version, SDP-DDPM fits the probability distribution and generates samples in SPD space. In the conditional version, the model generates the distribution of SPD given a specific condition and calculates the expectation $E(X | y)$ as the prediction. A high-capacity SPD U-Net structure is further introduced to improve the data fitting performance. Experimental results demonstrate the strong capability of the proposed method in the fitting probability distributions.
Acknowledgements

This paper is sponsored by the National Natural Science Foundation of China (NO. 62102151, NO. 12371289), the Shanghai Sailing Program (21YF1411200), Basic Research Project of Shanghai Science and Technology Commission(NO. 22JC1400800), CCF-Tencent Rhino-Bird Open Research Fund, the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science, Ministry of Education (KLATASDS2303), and the Fundamental Research Funds for the Central Universities.

References


