Gradient-based Riemannian Gaussian Differential Privacy

Rongpeng Yan [1, *], Mo Yu [2]

[1] Sports Engineering College, Beijing Sport University, Beijing, China, 100091

[2] Department of Management Science and Engineering, Harbin Institute of Technology, Harbin, China, 150006

* Corresponding Author Email: 18840697359@163.com

Abstract. Taking into account the non-linear structure of the Riemannian manifold, this paper proposes a Riemannian Gaussian differential privacy method based on gradient optimization. Specifically, the proposed method defines a Riemannian Gaussian distribution on the manifold by using the geodesic metric, and constructs a new sensitivity calculation method by using an exponential mapping. Furthermore, this method is the first to extend the gradient-based Euclidean Gaussian differential privacy technique to the Riemannian manifold, broadening the scope of differential privacy techniques. Simulation experiments were conducted on the unit sphere and the symmetric positive definite matrix manifold (SPD matrices), and the results show that the proposed method satisfies the differential privacy conditions, thereby demonstrating the effectiveness of the method.

Keywords: Gradient optimization, differential privacy, Gaussian differential privacy, Riemannian manifolds.

1. Introduction

With the rapid development of technology, we face new challenges when processing and analyzing complex data. These data are usually located in nonlinear manifolds and are widely used in fields such as medical imaging, financial investment, social networking, etc.[1-4]These nonlinear manifolds have unique geometric properties and can extract more valuable information from complex data. As the complexity of data increases, so does the importance of protecting data privacy. Differential privacy (DP) is the first widely recognized privacy protection theory, and its related mechanism aims to apply DP to real data processing. However, traditional differential privacy mechanisms, while performing well when dealing with linear data [5], often face challenges when dealing with complex non-linear data. To solve this problem, a common approach is to embed nonlinear data into Euclidean space and then apply standard DP mechanism for privacy protection. However, as Reimherr et al. (2021) point out, the inherent geometry of nonlinear data can enable higher data utility while preserving data privacy. Therefore, privacy mechanisms need to refer to the knowledge of differential geometry [6] and combine the unique geometry of the data to adapt to the complexity of nonlinear data.

Reimherr et al first introduced consideration of general manifolds in the differential privacy literature in 2021. They extend the Laplacian mechanism of ε-DP [7] from Euclidean space to a Riemannian manifold [8], and by studying the Frechette mean privatization problem, prove that optimal privacy protection and data utility can be achieved only when the underlying geometry of the data is utilized. In order to further improve this theory, Soto et al. developed a k-norm gradient mechanism for ε-DP on Riemannian manifolds in 2022, and showed that this mechanism performs better than the previous Laplacian mechanism in the Frechette mean privatization task through theoretical analysis and experimental verification. In the same period, Utpala et al. extended (ε, δ) -DP and its Gaussian mechanism in 2023 [9-10], although their work was limited to a specific manifold of symmetric positive definite matrix Spaces (SPDM) [11]. By combining an SPDM space with a logarithmic Euclidean metric, they turned it into a space of geometric planes, an approach that simplified the research process and reduced complexity, but also somewhat sacrificed the generality of the method. Unlike other studies that focus on privacy protection by publishing polymorphic values,

Han et al and Utpala et al in 2022 and 2023 focus on solving the empirical risk minimization problem in a (ε, δ) -DP-compatible manner by privatizing the gradient [12] located on the cutting bundle of a Riemannian manifold.

Although Reimherr et al. have extended ε-differential privacy and its Laplacian mechanism to general Riemannian manifes in 2021, other variants of differential privacy are equally noteworthy. For example, studies by Dwork (2006), Mironov (2017), Bun and Steinke (2016), Dong et al. (2022) show that these variants each have unique advantages compared to the traditional definition of DP. Therefore, these variants should also be included in the scope of extended research. As one of these variants, Gaussian differential privacy (GDP) is superior to ε-DP in terms of composition and subsampling characteristics. In addition, the study also shows that in the case of composition constraints, all hypotheses based privacy definitions tend to a guarantee of GDP. As the dimensions of the privatized data approach infinity, the private mechanism for massively adding noise turns out to be asymptotically GDP. These characteristics make GDP a key definition among the different variants of differential privacy and make it the most suitable choice for generalization to Riemannian manifold.

With the rapid development of data collection and storage technology, there are more and more data on Riemannian manifold with nonlinear structure. The above differential privacy methods are constructed based on Euclidean data, so if directly applied to Riemann manifold data, it may violate the inherent geometry of Riemann manifold. Based on this, the Riemann Gaussian difference privacy method based on gradient optimization is proposed for the unit sphere[13]and the symmetric positive definite matrix manifold (SPD matrix)[14].

2. The basic theory and method

2.1. Riemann Gaussian distribution

2.1.1. Definition of geodesic distance

On a sphere, the geodesic distance is the length of the great arc between two points p,q on the sphere, namely:

$$\text{distance} = \arccos < \text{p}, \text{q} > \tag{1}$$

On the SPD manifold, the geodesic distance is calculated based on logarithmic mapping and Frobenius norm:

$$\text{distance} = \| \log (m1^{-1} \cdot m2) \|_F \tag{2}$$

2.1.2. Exponential change and logarithmic change

For SPD manifolds, the exponential change is:

$$\exp_A(V) = A^{\frac{1}{2}} \cdot \exp (A^{-\frac{1}{2}} V A^{-\frac{1}{2}}) \cdot A^{\frac{1}{2}} \tag{3}$$

Where A is the SPD matrix and V is the tangent vector at A. Logarithmic change:

$$\log_A(B) = A^{-\frac{1}{2}} \cdot \log (A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) \cdot A^{-\frac{1}{2}} \tag{4}$$

Where A and B are the SPD matrix for a sphere, its exponential change:

$$\exp_p(v) = \cos (\| v \|) \cdot p + \sin (\| v \|) \cdot \frac{v}{\|v\|} \tag{5}$$

Where p is a point on the sphere and v is the tangent vector at point p and the norm of the tangent vector. Logarithmic change:

$$\log_p(q) = \frac{\theta}{\sin (\theta)} \cdot (q - \cos (\theta) \cdot p) \tag{6}$$

Where p and q are two points on the sphere θ=arccos<p, q>is the spherical distance between them.

2.1.3. Definition of sensitivity

On a sphere we define:

$$\text{Sensitivity} = \frac{|f(p)-f(q)|}{\text{dist}(p,q)} \tag{7}$$

Where $\text{dist}(p,q)$ is the geodesic distance between the points p and q, $|f(p)-f(q)|$ is the difference in the function values of these two points.

Similarly in SPD manifolds:

$$\text{Sensitivity} = \frac{|f(m1)-f(m2)|}{\text{dist}(m1,m2)} \tag{8}$$

Where $\text{dist}(m1,m2)$ is the geodesic distance between the matrices m1 and m2, $|f(m1)-f(m2)|$ is the difference in function values between the two matrices.

2.1.4. Riemann Gaussian distribution

In a traditional Euclidean space, the density function of a Gaussian distribution is defined as follows:

$$p(x) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{9}$$

Where,μ is the mean vector,Σ is the covariance matrix$|\boldsymbol{\Sigma}|$is the determinant.

The density function of the Gaussian distribution on the SPD manifold can be expressed as:

$$p(X) = \frac{1}{Z}exp\left(-\frac{1}{2}dist^2(X,\mu)\right) \tag{10}$$

Where $dist(X,\mu)$ represents the distance between the matrix X and the mean matrix μ on the SPD manifold, where Z is the normalization constant, ensuring that the integral of the density function is 1.

Specifically, if we use a logarithmic Euclidean metric, it can be calculated in the following way:

$$\text{dist}(X,\mu) = \| \log(X) - \log(\mu) \|_F \tag{11}$$

Here, $\|\cdot\|_F$denotes the Frobenius norm, $\log(X)$ represents taking the logarithm of the SPD matrix X.

Therefore, the Gaussian distribution density function on the SPD manifold has the following form:

$$p(X) = \frac{1}{Z}\exp\left(-\frac{1}{2}\| \log(X) - \log(\mu) \|_F^2\right) \tag{12}$$

Here X is a point on a manifold, specifically on the SPD (Symmetric Positive Definite) manifold, where X is a symmetric positive definite matrix.$\frac{1}{Z}$ Is the normalizing constant, which ensures that the product of the entire distribution is equal to 1?$\log(X)$And $\log(\mu)$ are the logarithm of the matrix X and μ.

2.2. Riemanngauss differential privacy based on gradient optimization

Firstly, the method of random number generation is used to generate data randomly on the manifold, and the sensitivity is taken as the optimization goal. The geodesic distance L between the samples and the function difference value Δy are calculated, and the sensitivity =Δy/L is defined. The gradient is estimated by making small perturbations to each element in the covariance matrix and observing changes in sensitivity. Based on the calculated gradient, the covariance matrix is updated using gradient descent to move it in the direction of reduced sensitivity**.** The eigenvalues of the updated covariance matrix need to be guaranteed to be non-negative to maintain symmetry and positivity. The standard definition of differential privacy is as follows:

$$Pr[\mathcal{M}(D_1) \in S] \leq e^{\epsilon}Pr[\mathcal{M}(D_2) \in S] + \delta \tag{13}$$

Where $\mathcal{M}$ is the differential privacy mechanism, input data set D_1 or D_2 D_1 and D_2 are adjacent data sets (that is, they differ in only one data point). S Is any subset of the output space? ϵ Is the privacy loss parameter, δ is the allowed probability of failure.

For Gaussian mechanisms, to ensure differential privacy, the standard deviation of added noise should be proportional to the sensitivity. Specifically, the Gaussian mechanism requires that the noise variance satisfy:

$$\sigma^2 \geq \frac{\Delta f^2 \cdot 2\ln(1.25/\delta)}{\epsilon^2} \tag{14}$$

For Gaussian mechanisms on Riemannian manifolds, we still add noise to guarantee difference privacy. To ensure differential privacy, it is necessary to prove that the optimized noise intensity satisfies the inequality in the definition of differential privacy:

$$\sigma_{\text{opt}}^2 \geq \frac{2\ln(1.25/\delta)\cdot\Delta f_{\text{opt}}^2}{\epsilon^2} \tag{15}$$

Δf_{opt} is the optimal sensitivity after gradient optimization. The goal of gradient optimization is to find the optimal covariance matrix that minimizes the sensitivity on the sphere or SPD manifold, thus minimizing the required noise intensity. The sensitivity calculation can be expressed as:

$$\Delta f_{\text{opt}} = \max_{D_1, D_2 \text{ are adjacent}} \| f(D_1) - f(D_2) \|_{g_{\Sigma_{\text{opt}}}} \tag{16}$$

Where, $\|\cdot\|_{g_{\Sigma_{\text{opt}}}}$ represents the geodesic distance defined under the covariance matrix Σ_{opt}.

Take advantage of optimized sensitivity Δf_{opt}, add noise intensity σ_{opt}^2 must meet:

$$\sigma_{\text{opt}}^2 = \frac{2\ln(1.25/\delta)\cdot\Delta f_{\text{opt}}^2}{\epsilon^2} \tag{17}$$

When we add Gaussian noise that satisfies the above variance, the output distribution generated by the mechanism $\mathcal{M}$ must satisfy the inequality (14). so the mechanism satisfies (ϵ, δ) - differential privacy. Specific applications are as Table 1-2.

Table 1. Application one: Gaussian differential privacy on a sphere.

Application one:Gaussian differential privacy on a sphere
1. Generate random samples to calculate sensitivity
2.Gradient calculation: Gradient matrix grad is set to store the gradient of each covariance matrix element; perturbation of perturbation covariance matrix perturbation=1e-4
3.Updated covariance matrix: learning_rate=0.01
4.Ensure the symmetry and positive quality of the covariance matrix: covariance=(covariance+covariance.T)/2; Minimum eigenvalue threshold of positive qualitative guarantee epsilon_for_pos_def=1e-6
5.Adjust the noise level to meet differential privacy requirements: differential privacy parameter epsilon (ϵ) = 0.5 delta (δ) = 1e-5

Table 2. Application two: Gaussian differential privacy on SPD.

Application two:Gaussian differential privacy on SPD
1The lower limit of noise design: noise_floor = 0.3
2.The differential privacy parameter epsilon (ϵ) = 0.5 delta (δ) = 1e-5
3.Design learning rate range: initial learning rate initial_learning_rate = 0.01, final learning rate final_learning_rate = 0.0001
4.L2 regularization weight 2.0 and L1 regularization weight 0.1 were used to control model complexity and parameter sparsity. reg_weight_L2 = 2.0 reg_weight_L1 = 0.1
5.Force the covariance matrix to be symmetric and positive: covariance=(covariance+covariance.T)/2; eigenvalues=np.clip(eigenvalues, noise_floor, None)

3. Results

3.1. Experimental data and environment

This paper uses python language to conduct simulation experiments. The experimental environment was Intel64 Family 6 Model 154 Stepping 3, GenuineIntel, 15.69GB memory, Windows-11-10.0.22631-SP0 64bit operating system. The data of simulation experiment were generated by random number generation method.

3.2. Analysis of experimental results

In this paper, two sets of experiments were carried out to achieve Gaussian difference privacy on spherical and SPD manifolds respectively, and the enhancement effect of the proposed optimization model was verified by comparing the relevant parameters of the model before and after optimization and the image comparison.

In this paper, we first show the before-and-after graphs after Gaussian differential privacy is realized on the spherical surface and optimized by using the gradient descent optimization model. The evaluation indexes were sensitivity change and variance change.

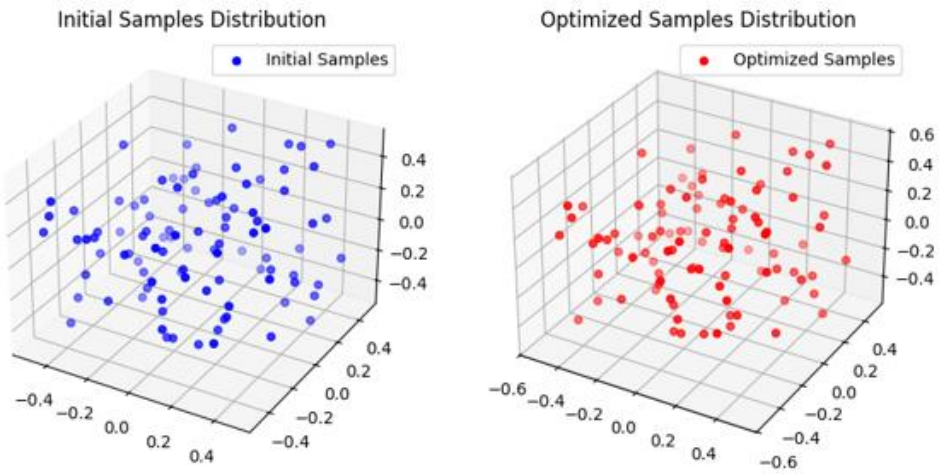


Figure 1. Comparison before and after spherical optimization.

The above figure is a before and after comparison of Gaussian differential privacy on a sphere by gradient descent optimization algorithm. As shown in the image, compared with the original model, the sample coverage of the optimized model has been significantly improved, while the dependence of the output results on the input changes is basically the same, indicating that the optimization algorithm has improved the degree of data privacy protection.

Table 3. Spherical result tables for different sample sizes.

Sample size	Primary sensitivity	Optimized sensitivity	Original variance	Optimized variance
10	0.914159	0.918712	0.084145	0.923815
30	0.979183	0.980943	0.98035	0.971561
50	0.997008	0.979096	0.986699	0.996533
80	0.992187	0.993320	0.997101	0.966046
100	0.994610	0.983932	0.950869	0.967396

It can be observed from Table 3-4 and Figure 1-4 that under different sample sizes, sensitivity generally shows an upward trend with the increase of sample size, but the sensitivity of the optimized model is slightly higher than that of the original model. In terms of variance, with the increase of sample size, the variance gradually tends to be stable, and the variance of the optimized model is significantly lower than that of the original model under most sample sizes. This shows that the optimized model can effectively reduce the variance while ensuring high sensitivity, thus improving the stability and accuracy of the model and achieving the effect of protecting the difference privacy.

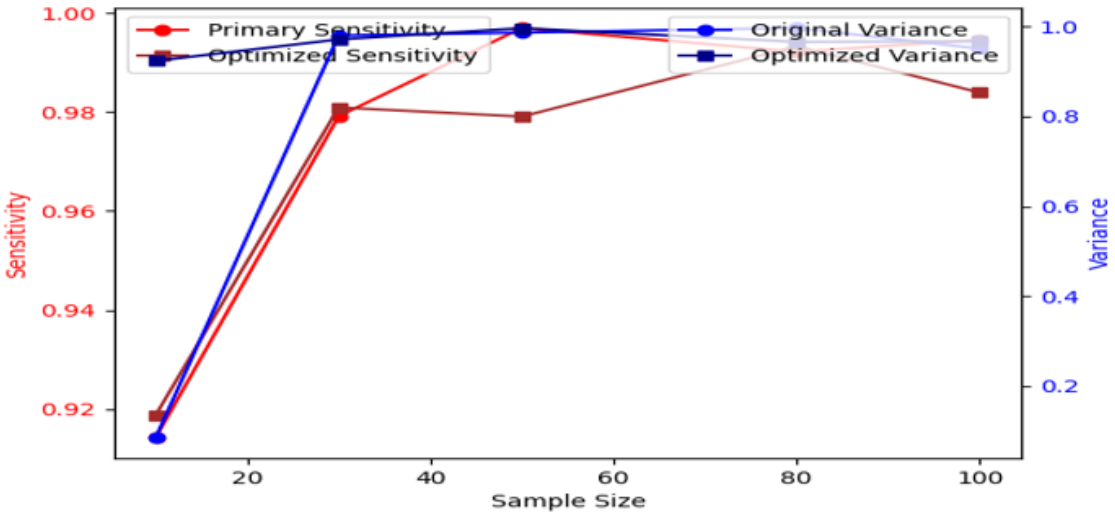


Figure 2. Spherical line plots with different sample sizes.

As can be seen from the scatter plot, the sample distribution before optimization is relatively concentrated, the coverage is small, and the distance between samples is relatively close, indicating that the diversity of initial samples is low. However, after optimization, the sample distribution is significantly more extensive, covering a larger spatial area, and the distribution is more uniform, reflecting that the adjustment of covariance matrix effectively increases the diversity of samples. After optimization, the sensitivity decreases, which makes the model's response to input disturbance more stable, the variance increases significantly, and the samples are more representative and diverse, which verifies the effectiveness of the optimization process.

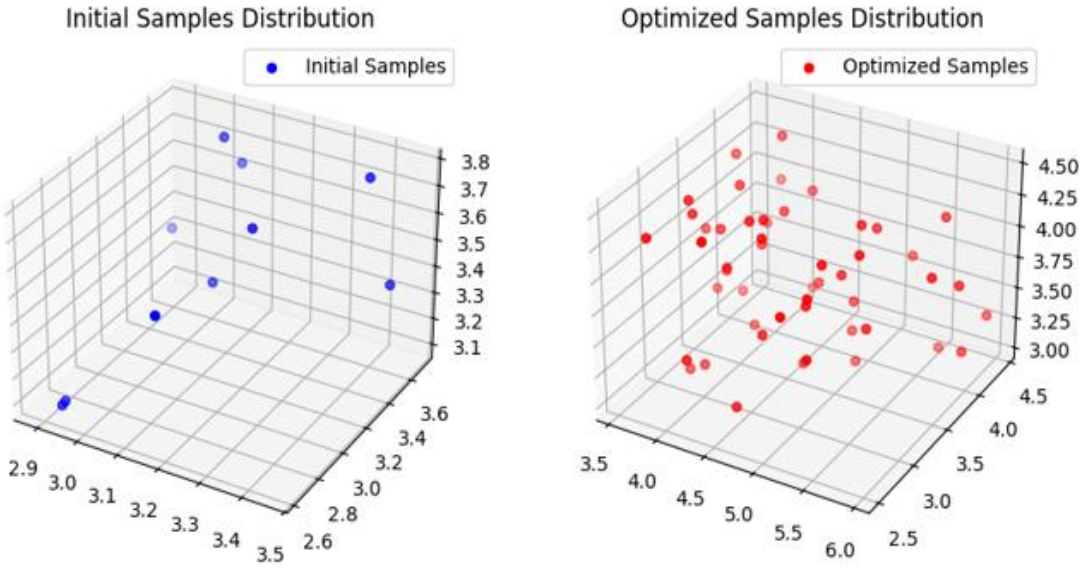


Figure 3. SPD manifold optimization before and after comparison.

According to the image analysis results, the optimized model generally shows higher sensitivity under different sample sizes, especially when the sample size is 30 and 50, the sensitivity of the optimized model is significantly improved. This shows that through the optimization process, the model becomes more perceptive in capturing data features and is able to reflect small changes in the data more effectively. At the same time, although the variance of the optimized model increases slightly in some cases, the increase in variance is relatively mild and does not significantly affect the stability of the model. Therefore, the optimization process succeeded in improving the sensitivity while keeping the variance relatively stable. This result shows that the performance of the model can be significantly enhanced by a reasonable optimization strategy without significantly increasing the volatility of the model, and the privacy protection can be effectively achieved.

Table 4. SPD manifold result table for different sample sizes.

Sample size	Primary sensitivity	Optimized sensitivity	Original variance	Optimized variance
10	6.10185	5.65780	0.02796	0.08272
30	6.76005	6.90296	0.03027	0.08743
50	6.86523	7.44158	0.03589	0.09504
80	6.89872	7.06970	0.03297	0.08804
100	6.88373	7.64484	0.03303	0.10864

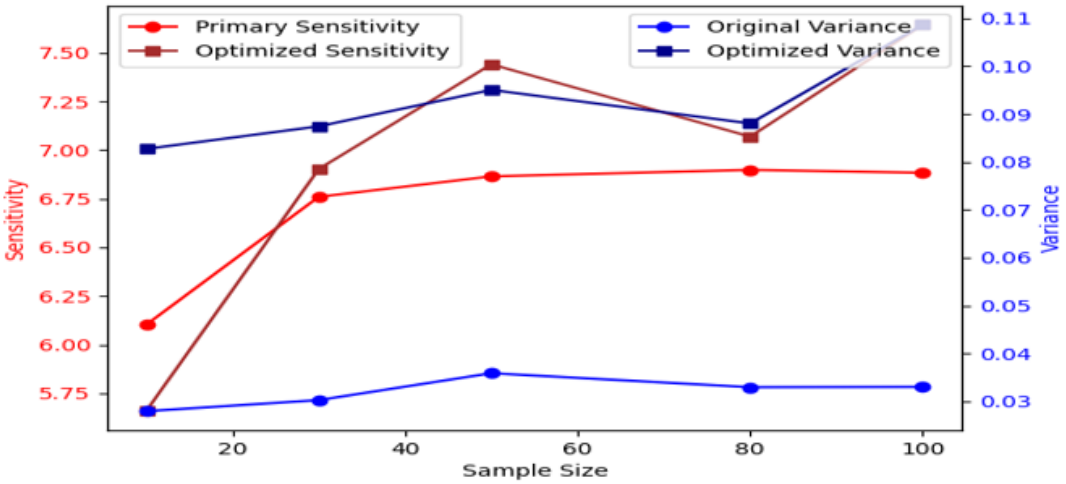


Figure 4. SPD manifold line diagram with different sample sizes.

4. Conclusions

In this experiment, we use the method of gradient descent combined with Gaussian differential privacy to optimize the model, and analyze the experimental results in detail. The experimental results show that although the sensitivity is slightly increased after optimization, the model is significantly improved in terms of the diversity and distribution of generated samples. Specifically, before optimization, the sample distribution was relatively concentrated and the coverage was small. However, after optimization, the sample distribution scope was significantly expanded, reflecting a significant increase in the difference between samples, and the diversity and generalization ability of the model were effectively improved. However, some shortcomings have also been exposed in the optimization process.

First, the increased sensitivity means that to a certain extent, the model is more sensitive to changes in the input data, which may have some impact on privacy protection. While increased sensitivity brings positive results in terms of model diversity and wide distribution, it also reminds us of the need to pay more attention to the balance between privacy protection and model stability during optimization. For this reason, future studies may try to introduce a more refined sensitivity control mechanism to further reduce sensitivity by dynamically adjusting the sensitivity threshold, so as to minimize the dependence on sensitive data while maintaining the diversity and generalization ability of the model, so as to enhance the privacy protection effect.

In addition, in terms of noise addition, although the current Gaussian noise model effectively increases the diversity of samples, excessive noise addition may have a negative impact on model performance. Future research could explore the applicability of different types of noise models (such as Laplacian noise, Gaussian noise, etc.) in different application scenarios and optimize the way noise is added to minimize its negative impact on model performance. In addition, extending the current method to more complex data structures and high-dimensional data application scenarios is also an important direction for future research. This includes more complex data types such as graph data and time series data to verify the effectiveness and applicability of the method in the field of big data

analysis and deep learning. With these improvements, we hope to extend this optimization approach to a variety of practical applications, such as healthcare, finance, and government statistics, and ultimately find a better balance between protecting privacy and improving model performance to provide stronger support for data-driven decision making.

References

[1] Wang X, Cui W, Wu H, et al. Hybrid-feature based spherical quasi-conformal registration for AD-induced hippocampal surface morphological changes [J]. Computer Methods and Programs in Biomedicine, 2024,256.

[2] Regulation S-P: Privacy of Consumer Financial Information and Safeguarding Customer Information; Correction [J]. The Federal Register / FIND, 2024, 89(124).

[3] Gao Rui, Chen Xuebin, Gu Zheng, et al. Satisfy personalized social network graph generating methods of differential privacy [J]. Journal of taiyuan university of technology, 2024 zhongguo kuangye daxue (01): 163-171. The DOI: 10.16355 / j.t. yut. 1007-9432.2023 BD001.

[4] Agarwal R, Kumar M. A privacy preserving scheme for encrypted medical images using RDH [J]. International Journal of Information and Computer Security, 2023, 22(3-4).

[5] Hu H, Liu R, Liu R. An intelligent diagnostic model for industrial equipment with privacy protection [J]. Computers & Security, 2024,146.

[6] S. M S. On the Solvability of Linear Differential Operators on Vector Bundles over a Manifold [J]. Differential Equations, 2024, 59(12).

[7] G. K S ,K. P ,G. M U , et al. Differential privacy scheme using Laplace mechanism and statistical method computation in deep neural network for privacy preservation[J]. Engineering Applications of Artificial Intelligence, 2024,128.

[8] Yan Yao-Yao. Gradient estimation of a class of quasi-linear elliptic equations with Neumann boundary value conditions on Riemannian manifolds [D]. Shanghai Normal University, 2024.

[9] Li Min, Xiao Di, Chen Lujun. Adaptive Gaussian Differential Privacy Personalized Federated Learning considering communication efficiency and utility [J]. Journal of Computer Science, 2013, 47(04):924-946.

[10] Yang Yiqi, Zhang Bangning, Guo Daosheng, et al. Data freshness performance analysis for Gaussian Differential privacy in swarm intelligence sensing systems [J]. Radio Engineering, 2019, 54(03):526-534.

[11] Wang X, Diao Y, Han D. Does digital industry agglomeration bring substantive green innovation?—Empirical analysis of SPDM and PTR models based on spatiotemporal dual fixation [J]. Managerial and Decision Economics, 2023, 45(3).

[12] Raju K S S ,Durgaprasad P ,Palencia D L J , et al. Contour Analysis for Heat Transfer Rate in a Wedge Geometry with Non-Uniform Shapes Nanofluid: Gradient Descent Machine Learning Technique[J]. Results in Engineering, 2024, 23.

[13] Reinhard S ,Raghed K ,Nichole O , et al. Conformational Manifold Sampled by Two Short Linear Motif Segments Probed by Circular Dichroism, Vibrational, and Nuclear Magnetic Resonance Spectroscopy.[J]. Biochemistry, 2023, 62(17).

[14] Chen J, Tang J, Shao W, et al. A new numerical simulation method of 3D rough surface topography with coupling 3D roughness parameters Sdr, Sdq, Spd, Spc, and characteristic functions [J]. Tribology International, 2024,200.