Advances in Applied Mathematics 应用数学进展, 2025, 14(2), 162-170
Published Online February 2025 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2025.142061

Hans汉斯

一种基于差分隐私的矩阵填充算法

郭佳浩

温州大学数理学院，浙江 温州

收稿日期：2025年1月18日；录用日期：2025年2月11日；发布日期：2025年2月20日

摘　要

矩阵填充作为矩阵分析中的一个关键课题，对于处理不完整数据集至关重要，其应用广泛，覆盖图像处理、推荐系统等领域。与此同时，隐私保护在数据安全领域占据核心地位，旨在确保个人敏感信息在数据分析过程中得到充分的安全保障和保密处理。将这两个领域的研究结合，为应对实际应用中的复杂挑战提供了创新性的解决方案。文章提出了一种基于差分隐私的新型隐私保护矩阵填充方案。该方案巧妙地融合了差分隐私理论框架，通过向原始数据中添加噪声来保护用户隐私。这种做法确保了即使在数据公开或共享的情况下，个体的信息也不会被泄露，从而有效地防止了潜在的隐私风险。数值实验结果表明，所提出的方案不仅能够显著保护用户的隐私，同时还在保持数据效用方面表现出色。此外，该方法还提升了算法运行效率，降低了计算成本。

关键词

矩阵填充，差分隐私，交替方向乘子法

A Matrix Completion Algorithm Based on Differential Privacy

Jiahao Guo

College of Mathematics and Physics, Wenzhou University, Wenzhou Zhejiang

Received: Jan. 18th, 2025; accepted: Feb. 11th, 2025; published: Feb. 20th, 2025

Abstract

Matrix completion, as a key topic in matrix analysis, is crucial for dealing with incomplete datasets, and its applications cover a wide range of fields, such as image processing and recommender systems. Meanwhile, privacy protection occupies a central position in the field of data security, aiming to ensure that personal sensitive information is adequately secured and confidentially handled during data analysis. Combining research in these two fields provides innovative solutions to address

文章引用: 郭佳浩. 一种基于差分隐私的矩阵填充算法[J]. 应用数学进展, 2025, 14(2): 162-170.
DOI: 10.12677/aam.2025.142061

the complex challenges in practical applications. In this paper, we propose a novel privacy-preserving matrix completion scheme based on differential privacy. The scheme cleverly incorporates a differential privacy theoretical framework to protect user privacy by adding noise to the raw data. This approach ensures that individuals' information is not disclosed even when the data is public or shared, thus effectively preventing potential privacy risks. Numerical experimental results show that the proposed scheme not only significantly protects user privacy, but also excels in maintaining data utility. In addition, the method improves the algorithm operation efficiency and reduces the computational cost.

Keywords

Matrix Completion, Differential Privacy, Alternating Direction Method of Multipliers

1. 介绍

矩阵填充[1]-[3]，也称为矩阵补全，是一种旨在从不完整或受损的数据矩阵中准确且高效地重建原始数据的技术。在处理大规模数据集时，经常会遇到数据元素的缺失、污染和损坏等问题，这些问题可能导致传统分析方法的失效。因此，研究如何精确且有效地恢复这些数据中的缺失或损毁部分具有重要的现实意义。矩阵填充技术利用了原始数据潜在的低秩特性来估计和填补缺失值，这使得它成为一种关键工具，并广泛应用于多个领域，如图像修复，即通过矩阵恢复技术来修补被遮挡或损坏的图像部分[4]。当用户将这类数据上传至云端进行处理时，由于数据的“易复制性”，如果以未加密的形式传输，可能会面临较高的泄漏风险[5]。一旦数据泄露，其价值可能显著降低。差分隐私[6] [7]是一种用于保护个人数据隐私的先进方法，它通过在查询结果中引入适量的随机噪声，来防止攻击者从发布的数据或分析结果中识别出任何特定个体的信息。此方法确保了即使在大规模数据集上进行复杂的分析，个人隐私也能得到有效保护。这种方法已被广泛应用于数据统计和机器学习领域，以确保数据分析过程中的隐私安全。本文提出了一种创新性的解决方案，结合差分隐私机制，设计出一种高效的隐私保护矩阵填充方法。

2. 矩阵填充

恢复缺失值的不完整矩阵问题最近引起了图像处理[8] [9]、信号处理[10] [11]和机器学习[12] [13]领域研究人员的极大关注。传统方法将这一任务表述为低秩矩阵最小化问题。假设 $M \in R^{m\times n}$ 是一个不完整矩阵，那么，传统的低秩矩阵最小化问题表述如下：

$$\min_{X} \operatorname{rank}(X), \quad \text{s.t. } X_{i,j} = M_{i,j},\ i,j \in \Omega \tag{1}$$

其中 $X \in R^{m\times n}$ 是已知的低秩矩阵，$\operatorname{rank}(X)$ 表示的是矩阵 X 的秩，Ω 是与观察到的条目对应的位置集。

式(1)中的问题是 NP 难问题，但在广泛的条件[14]下，该问题可以通过核范数最小化实现优化。因此式(1)可以改写成：

$$\min_{X} \|X\|_* \quad \text{s.t. } X_{i,j} = M_{i,j},\ i,j \in \Omega \tag{2}$$

其中 $\|X\|_*$ 表示矩阵 X 的奇异值之和。在矩阵恢复的研究中，核范数作为凸优化问题中的一个关键元素，最初被广泛应用于低秩矩阵填充问题。然而，随着研究的深入，研究人员发现单纯依赖核范数进行矩阵

恢复存在一定的局限性，尤其是在准确率方面未能达到预期效果。因此，为了克服这一挑战，学术界和工业界开始探索其他类型的范数来改进矩阵填充的效果。例如，Schatten p-范数[15]作为一种非凸范数，在一定程度上提供了比传统核范数更好地逼近低秩结构的能力，从而提高了矩阵填充的准确性。加权核范数[4]则通过引入权重参数，允许对不同奇异值施加不同程度的影响，进一步增强了模型的灵活性和适应性。这些新提出的范数确实提升了矩阵填充的准确度，但它们仍然依赖于 SVD 来获取不完全矩阵的奇异值。尽管 SVD 是处理矩阵分解的经典方法，但它计算复杂度较高，尤其是在处理大规模矩阵时，其速度往往不能满足实际应用的需求。为了提高矩阵分解的速度，研究者们提出了多种基于矩阵分解的新方法。2013 年，一种创新的方法——快速三元分解(Fast Tri-Factorization, FTF)方法[16]被提出，该方法基于 Qatar Riyal (QR)分解[17]而非传统的 SVD 分解。QR 分解是一种数值稳定性较好的矩阵分解方式，它将一个矩阵分解成一个正交矩阵 Q 和一个上三角矩阵 R 的乘积。FTF 方法利用了 QR 分解的优势，不仅能够有效降低计算复杂度，还能保证较高的分解精度，从而显著提高了矩阵分解的速度，为大尺度矩阵填充问题提供了一种更为高效的技术手段。2018 年，Liu 等人提出了一种 CSVD-QR 分解的方法，该方法大大提高了矩阵分解的速度。该算法的主要步骤如算法 1 所示：

算法 1. CSVD-QR 算法

输入：$X \in \mathbb{R}^{m\times n}$，目标秩 $r = 0.01\times \min(m,n)$，迭代次数 t，设定三个初始矩阵 $L_0 = eye(m,r)$，$R_0 = eye(r,n)$，$D_0 = eye(r,r)$

输出：生成矩阵 $X = LDR$

1. **while** $\|L_k * D_k * R_k - X\|_F^2 \geq \varepsilon_0$ 和 $k \leq Iter$ **do**
2. 计算 QR 分解：$XR_{k-1}^{\mathrm{T}} = L_k T$
3. 计算 QR 分解：$X^{\mathrm{T}} L_k = R_k D_k$ 并且生成 $R_k = R_k^{\mathrm{T}}$
4. **end while**
5. 令 $L = L_k$，$D = D_k^{\mathrm{T}}$，$R = R_k$

首先，随机生成三个单位矩阵 $L_0 = eye(m,r)$，$R_0 = eye(r,n)$，$D_0 = eye(r,r)$，其中 $r = 0.01\times \min(m,n)$，算法的第 k 次迭代过程如下：

第一步，固定 D_{k-1} 和 R_{k-1} 来更新 L_k，基于矩阵 R_{k-1} 是正定的，因此有 $L_k D_{k-1} = XR_{k-1}^{\mathrm{T}}$，该式子可以写成 $[L_k, \sim] = qr\left(XR_{k-1}^{\mathrm{T}}\right)$；

第二步，固定 L_{k-1} 和 D_{k-1} 来更新 R_k，基于矩阵 L_{k-1} 是正定的，有 $\left[R_k^{\mathrm{T}}, D_{k-1}^{\mathrm{T}}\right] = qr\left(X^{\mathrm{T}} L_k\right)$；

第三步，固定 L_{k-1} 和 R_{k-1} 来更新 D_k，同上可以得到 $\left[R_k^{\mathrm{T}}, D_k^{\mathrm{T}}\right] = qr\left(X^{\mathrm{T}} L_k\right)$。

3. 差分隐私之拉普拉斯机制

定义 1：假设有随机算法 $\mathcal{M}$，S 为 $\mathcal{M}$ 所有可能输出结果构成的集合，$Pr[.]$ 表示概率，对于任意两个相邻数据集 D、B，两个数据集的差别只有 1 条记录，如果满足：

$$Pr\left[\mathcal{M}(D) \in S\right] \leq \exp^{\epsilon} Pr\left[\mathcal{M}(B) \in S\right] + \delta \tag{3}$$

则称算法 $\mathcal{M}$ 提供 ϵ-差分隐私保护，其中 ϵ 为差分隐私预算，用来保证数据集中增加或减少一条记录，随机算法 $\mathcal{M}$ 的输出结果一致的概率。ϵ 越接近 0，$\mathcal{M}$ 在 D、B 上输出的数据分布越接近，输出结果越不可区分，隐私保护程度越高。δ 是用于限制模型行为任意改变的概率，通常设置为一个小的常数，推荐设置小于训练数据集大小的倒数。

为了实现差分隐私，最常用的方法是加入噪声。在本文中，我们选择对观测矩阵进行拉普拉斯噪声的方法来实现保护用户隐私的目的。下面给出拉普拉斯噪声的定义。

定义 2：拉普拉斯噪声是指从拉普拉斯分布中抽取的随机变量。拉普拉斯分布是一种连续型概率分布，其概率密度函数为：

$$p\left(x\middle|\mu,b\right)=\frac{1}{2b}\exp\left(-\frac{\left|x-\mu\right|}{b}\right) \tag{4}$$

其中 μ 是位置参数，它决定了分布的中心位置，b 是尺度参数，它决定了分布的宽度。由概率密度函数求分布的概率累计函数如下：

$$\mathrm{F}\left(x\middle|\mu,b\right)=\begin{cases}\dfrac{1}{2}\exp\left(-\dfrac{\mu-x}{b}\right), x<\mu\\ 1-\dfrac{1}{2}\exp\left(-\dfrac{\mu-x}{b}\right), x\geq\mu\end{cases} \tag{5}$$

推导过程如下：

当 $x<\mu$ 时，$f\left(x\middle|\mu,b\right)=\dfrac{1}{2b}\exp\left(-\dfrac{\mu-x}{b}\right)$：

$$\mathrm{F}\left(x\middle|\mu,b\right)=\frac{1}{2b}\int_{-\infty}^{x}\exp\left(-\frac{\mu-x}{b}\right)\mathrm{d}x=\frac{1}{2b}\int_{-\infty}^{x}\exp\left(\frac{x-\mu}{b}\right)\mathrm{d}x \tag{6}$$

令 $t=\dfrac{x-\mu}{b}$，可求得：

$$\mathrm{F}\left(x\middle|\mu,b\right)=\frac{1}{2b}\int_{-\infty}^{\frac{x-\mu}{b}}b\mathrm{e}^{t}\mathrm{d}t=\frac{1}{2}\int_{-\infty}^{\frac{x-\mu}{b}}\mathrm{e}^{t}\mathrm{d}t=\frac{1}{2}\left[\mathrm{e}^{t}\right]_{-\infty}^{\frac{x-\mu}{b}}=\frac{1}{2}\mathrm{e}^{-\frac{\mu-x}{b}} \tag{7}$$

那么当 $x\geq\mu$ 时，由于拉普拉斯噪声满足对称性，因此有以下等式成立：

$$\mathrm{F}\left(x\middle|\mu,b\right)=\frac{1}{2b}\int_{-\infty}^{x}\mathrm{e}^{-\frac{\mu-x}{b}}\mathrm{d}x=1-\frac{1}{2b}\int_{x}^{+\infty}\mathrm{e}^{-\frac{\mu-x}{b}}\mathrm{d}x=1-\frac{1}{2}\mathrm{e}^{-\frac{\mu-x}{b}} \tag{8}$$

下面我们来证明拉普拉斯噪声满足差分隐私。

给定一个映射函数 $f:D\to R^{d}$，它表示数据集 D 到一个 d 维空间的映射关系。我们在所得到的函数 $f\left(D\right)=\left(x_1,x_2,\cdots,x_d\right)^{\mathrm{T}}$ 上加上拉普拉斯噪声，得到一个输出函数 $\mathcal{M}\left(D\right)$。

那么有：

$$\mathcal{M}\left(D\right)=f\left(D\right)+\left(\mathrm{Lap}_1\left(\frac{\Delta f}{\varepsilon}\right),\mathrm{Lap}_2\left(\frac{\Delta f}{\varepsilon}\right),\cdots,\mathrm{Lap}_d\left(\frac{\Delta f}{\varepsilon}\right)\right)^{\mathrm{T}} \tag{9}$$

其中 $\Delta f=\max\limits_{D,B}\left\|f\left(D\right)-f\left(B\right)\right\|_p$，其中 p 一般取 1，为一范数。

下面，我们将证明 $\mathcal{M}\left(D\right)$ 满足差分隐私定义。

设 $f\left(D\right)=\left(x_1,x_2,\cdots,x_d\right)^{\mathrm{T}}$，$f\left(B\right)=\left(x_1',x_2',\cdots,x_d'\right)^{\mathrm{T}}=\left(x_1+\Delta x_1,x_2+\Delta x_2,\cdots x_d+\Delta x_d\right)^{\mathrm{T}}$。

$$\Delta f=\max\left(\sum_{i=1}^{d}\left|x_i-x_i'\right|\right)=\max\left(\sum_{i=1}^{d}\left|\Delta x_i\right|\right) \tag{10}$$

不失一般性，我们可以设 x_i 全部为 0。$f\left(D\right)=\left(0,0,\cdots,0\right)^{\mathrm{T}}$，$f\left(B\right)=\left(\Delta x_1,\Delta x_2,\cdots,\Delta x_d\right)^{\mathrm{T}}$。

记输出向量 $O=\left(y_1,y_2,\cdots,y_d\right)^{\mathrm{T}}$。

那么有

$$Pr\left[\mathcal{M}(D)=O\right]=\prod_{i=1}^{d}\frac{\varepsilon}{2\Delta f}\mathrm{e}^{-\frac{\varepsilon}{\Delta f}|y_i|} \tag{11}$$

$$Pr\left[\mathcal{M}(B)=O\right]=\prod_{i=1}^{d}\frac{\varepsilon}{2\Delta f}\mathrm{e}^{-\frac{\varepsilon}{\Delta f}|x_i-y_i|} \tag{12}$$

$$\frac{Pr\left[\mathcal{M}(D)=O\right]}{Pr\left[\mathcal{M}(B)=O\right]}=\frac{\prod_{i=1}^{d}\frac{\varepsilon}{2\Delta f}\mathrm{e}^{-\frac{\varepsilon}{\Delta f}|y_i|}}{\prod_{i=1}^{d}\frac{\varepsilon}{2\Delta f}\mathrm{e}^{-\frac{\varepsilon}{\Delta f}|x_i-y_i|}}=\prod_{i=1}^{d}\mathrm{e}^{-\frac{\varepsilon}{\Delta f}\left(|y_i|-|\Delta x_i-y_i|\right)}=\mathrm{e}^{\frac{\varepsilon}{\Delta f}\sum_{i=1}^{d}\left(|\Delta x_i-y_i|-|y_i|\right)} \tag{13}$$

现在呢，我们只要论证 $\sum_{i=1}^{d}\left(|\Delta x_i-y_i|-|y_i|\right)\le\Delta f$ 成立，就有 $\frac{Pr\left[\mathcal{M}(D)=O\right]}{Pr\left[\mathcal{M}(B)=O\right]}\le\mathrm{e}^{\varepsilon}$ 该式等价于式(3)。

对于每一个 $|\Delta x_i-y_i|-|y_i|$，其中，y_i 看成变量，由绝对值不等式可得：

$$-|\Delta x_i|\le|\Delta x_i-y_i|-|y_i|\le|\Delta x_i|$$

那么有：

$$\sum_{i=1}^{d}\left(|\Delta x_i-y_i|-|y_i|\right)\le\sum_{i=1}^{d}|\Delta x_i|\le\max_{D,B}\left(\sum_{i=1}^{d}|\Delta x_i|\right)=\Delta f \tag{14}$$

于是有：

$$\frac{Pr\left[\mathcal{M}(D)=O\right]}{Pr\left[\mathcal{M}(B)=O\right]}\le\mathrm{e}^{\varepsilon} \tag{15}$$

拉普拉斯噪声满足 ε-差分隐私得证。

由于数据集 D,B 具有对称性。因此 $Pr\left[\mathcal{M}(D)=O\right]\le\mathrm{e}^{\varepsilon}Pr\left[\mathcal{M}(B)=O\right]$ 成立的同时，$Pr\left[\mathcal{M}(B)=O\right]\le\mathrm{e}^{\varepsilon}Pr\left[\mathcal{M}(D)=O\right]$ 也成立。证毕。

下面我们给出拉普拉斯噪声生成的过程。第一步，我们需要先确定尺度参数 β，其中 $\beta=\Delta f/\epsilon$；第二步，从均匀分布中生成随机数 $\mu=rand(m,n)$；第三步，计算噪声，公式为 $L\_noise=\beta*\left(rand(m,n)-0.5\right)*\left(\log\left(1/rand(m,n)\right)\right)$。

此外，在实际应用场景中，选择合适的隐私参数 ϵ 需要综合考量多方面因素：对于高度敏感的数据，通常会选择较小的 ϵ 值以确保更强的隐私保护；而当数据主要用于统计分析而非精确查询时，则可以适当放宽对 ϵ 的限制，以保持数据的可用性和分析结果的准确性。此外，用户的隐私偏好也至关重要，可以通过用户调查或设置个性化选项来确定一个能够平衡隐私与效用的 ϵ 值。最后，通过实验验证不同 ϵ 值下的数据准确性和隐私泄露风险，可以帮助进一步优化 ϵ 的选择，确保既满足隐私保护的需求，又不影响数据的实用价值。ϵ 的选择是基于差分隐私的理论框架，目的是在隐私保护和数据效用之间找到平衡。

4. 一种新的差分隐私矩阵填充算法

在本节中，我们将提出一种新的基于差分隐私的矩阵填充算法 DLNM_QR。

给定一个真实矩阵 $X\in\mathbb{R}^{m\times n}$，对其添加拉普拉斯噪声得到

$$X\_noisy=X+L\_noise \tag{16}$$

随机生成一个索引矩阵 $\Omega_{i,j}$，定义如下：

$$\Omega_{i,j}=\begin{cases}1, & \text{其他}\\ 0, & \text{如果 } X_{i,j}\text{ 缺失}\end{cases} \tag{17}$$

乘积得到 $\tilde{X}=X\_noisy*\Omega_{i,j}$，由此我们得到加密后的缺失矩阵 $\tilde{X}$。

从[17]可知，$L_{2,1}$ 范数被成功应用于低秩矩阵填充中，因此问题(2)可以被表示为：

$$\min_{\tilde{X}}\frac{1}{\mu}\left\|\tilde{X}\right\|_{2,1}+\frac{1}{2}\left\|\tilde{X}-Y\right\|_{\mathrm{F}}^{2} \tag{18}$$

其中 $Y\in\mathbb{R}^{m\times n}$ 是一个给定的真实矩阵并且 $\mu>0$。对 $\tilde{X}$ 做 CSVD-QR 分解，得到 $\tilde{X}=WZP$，因为 W 和 P 和正交的，问题(18)可以被写成：

$$\min_{Z}\frac{1}{\mu}\left\|Z\right\|_{2,1}+\frac{1}{2}\left\|Z-E\right\|_{\mathrm{F}}^{2}\quad \text{s.t. } \tilde{X}=WZP,\ E=W^{\mathrm{T}}YP^{\mathrm{T}} \tag{19}$$

对于问题(19)，它的收缩算子表达式如下：

$$Z(:,j)=\frac{\left(\left\|Z(:,j)\right\|_{\mathrm{F}}-\frac{1}{\mu}\right)_{+}}{\left\|Z(:,j)\right\|_{\mathrm{F}}} \tag{20}$$

其中 $(x)_{+}=\max\{x,0\}$，$x\in(-\infty,+\infty)$ 是一个实数。

最后，通过固定 W_k,Z_k,P_k，我们可以按照如下公式更新 $\widetilde{X_{k+1}}$，Y_{k+1}：

$$\widetilde{X_{k+1}}=W_kZ_kP_k-\Omega\circ W_kZ_kP_k+\tilde{X} \tag{21}$$

$$Y_{k+1}=Y_k+\mu_k\left(X_{k+1}-W_kZ_kP_k\right) \tag{22}$$

$$\mu_{k+1}=\rho\mu_k \tag{23}$$

其中 $\rho\geq1$。这种将交替方向乘子法(Alternating Direction Method of Multipliers, ADMM)与 CSVD-QR 算法结合在一起的方法称为 DLNM_QR 算法。DLNM_QR 矩阵补全方法是一种专门用于解决矩阵中缺失值估计问题的算法。在 DLNM_QR 方法中，为了对缺失的数据进行补全，首先构建了一个目标函数，这个函数通常是关于矩阵低秩性质的一种度量。由于此目标函数是凸的，因此保证了任何局部最优解同时也是全局最优解，从而使得通过梯度搜索方法找到的解具有理论上的优越性。这种方法不仅加速了收敛速度，还提高了算法的可扩展性和灵活性，使其能够适应不同大小和类型的矩阵填充任务。

5. 数值实验

在本节中，利用合成数据集来验证算法的各项性能，并将本研究的方法与现有的方法进行比较。从计算速度和误差来验证本研究方法的可行性以及有效性。用相对标准误差(*RSE*)来对比收敛精度，*RSE* 定义如下：

$$RSE=\frac{\left\|X-Y\right\|_{F}}{\left\|Y\right\|_{F}}$$

其中 X,Y 分别代表恢复后的矩阵以及真实矩阵。

为了探究矩阵大小对算法的影响，固定观测率 $p=15\%$，矩阵秩 $r=10$，迭代次数 $Iter=400$，拉普拉斯噪声尺度参数 $\beta=1$，专注于考察不同规模矩阵对三种算法的影响。具体而言，本研究逐步增大矩阵的尺寸，从 200 × 200 到 2000 × 2000，以全面评估这些变化如何影响算法的表现。表 1 详细记录并对比了这三种算法在不同矩阵大小下的运行速度及其标准误差。由分析结果可得，提出的 DLNM_QR 算法在标准误差方面没有显著优于其他两种算法，但在运行效率上却展现出了明显的优势，尤其当处理大型矩阵时，这种优势变得更加突出。例如，在处理 2000×2000 的矩阵中，DLNM_QR 算法的运行速度相较于其他两种算法有了近乎 22 倍的提升。为了研究拉普拉斯噪声尺度参数 β 对算法的影响，我们固定其余的参

数，表 2 记录了随着 β 变化标准误差变化的情况。由结果可知，ϵ 越大，标准误差越小，虽然对比其他算法来说，在标准误差上并未有太大的提升，但是在速度上的提升是显著的，反映了此算法的有效性能。随着 ϵ 的增大，标准误差呈现出减小的趋势。这一现象表明，在适当的范围内增加 ϵ 值能够有效降低恢复矩阵与真实矩阵之间的差异，从而提升恢复精度。值得注意的是，尽管与其他现有的算法相比，本方法在减少标准误差方面并没有显著超越，但在计算效率上却表现出明显的优势。这种性能上的优化不仅反映了算法的有效性，同时也证明了算法隐私保护的有效性。

上述实验结果表明，DLNM_QR 算法在保持计算精度的同时实现了显著的速度增益，说明它在算法设计上的优化策略是成功的，这也为进一步研究提供了有价值的启示。通过对比实验，观察到 DLNM_QR 算法在处理低秩矩阵补全问题时，能够在确保数据重建误差处于可接受范围内的前提下，大幅度缩短计算时间。这种速度的提升并非以牺牲精度为代价，而是通过对算法内部机制的精巧设计与优化来实现的。下面通过计算复杂度来阐述一下，本文所提出的 DLNM_QR 算法主要依靠的是 QR 分解，QR 分解的计算复杂度是 $O\left(r^2\left(m+n\right)\right)$，该算法执行两次 QR 分解。因此计算复杂度为 $2O\left(r^2\left(m+n\right)\right)$；非线性 SOR 算法的计算复杂度为 $O\left(r^3+mnr+mn\right)$。对比两种算法的计算复杂度，显然 DLNM_QR 算法计算复杂度比较低，因此计算时间更短。

Table 1. Numerical results of the impact of matrix size on complete-in
表 1. 矩阵大小对填充影响的数值结果

矩阵大小	观测值 p	矩阵秩 r	算法	运行速度/s	迭代次数	标准误差
200 × 200	0.15	10	SOR	1.134501	400	0.69485
			DLNM_QR	0.129055	400	0.1992
			AM	~	400	不收敛
500 × 500	0.15	10	SOR	8.005175	400	0.089262
			DLNM_QR	0.979278	400	0.089262
			AM	321.898549	400	0.078076
1000 × 1000	0.15	10	SOR	41.163018	400	0.069397
			DLNM_QR	5.435061	400	0.069397
			AM	3088.747391	400	0.050587
2000 × 2000	0.15	10	SOR	395.808449	400	0.060423
			DLNM_QR	17.391248	400	0.060423
			AM	21727.373228	400	0.034986

Table 2. Numerical results of the impact of Laplacian noise parameter scale on complete-in
表 2. 拉普拉斯噪声参数尺度对填充影响的数值结果

矩阵大小	观测值 p	参数尺度 ϵ	算法	运行速度/s	迭代次数	标准误差
1000 × 1000	0.15	0.2	SOR	37.818816	400	0.35251
			DLNM_QR	17.604057	400	0.30315
			AM	2917.141632	400	0.12657
1000 × 1000	0.15	0.5	SOR	37.885889	400	0.13909
			DLNM_QR	17.901269	400	0.12089
			AM	3039.961031	400	0.050314

续表

1000 × 1000	0.15	1	SOR	41.163018	400	0.069397
			DLNM_QR	5.435061	400	0.071362
			AM	3088.747391	400	0.050587
1000 × 1000	0.15	2	SOR	37.318582	400	0.034677
			DLNM_QR	19.266138	400	0.030209
			AM	2984.514562	400	0.012561
1000 × 1000	0.15	5	SOR	38.225119	400	0.013868
			DLNM_QR	17.781655	400	0.012083
			AM	2861.221987	400	0.0050237
1000 × 1000	0.15	10	SOR	38.148592	400	0.0069336
			DLNM_QR	18.052101	400	0.0060417
			AM	3074.358845	400	0.0025118

6. 结论

本文提出了一套基于拉普拉斯噪声添加机制的矩阵填充算法。与传统的矩阵补全方案相比，所提出的算法不仅有效保障了用户数据的隐私，还大幅提高了算法的执行效率。这种提升主要来源于两方面：一方面，通过矩阵分解技术减少了计算复杂度；另一方面，在不牺牲隐私保护强度的情况下最小化了额外的计算开销。未来，将会考虑将该算法应用到实际场景中，如图片恢复、轨迹恢复以及推荐系统。例如，在推荐系统中，它可以实现考虑数据隐私的协同过滤，使得个性化推荐既准确又安全，不会泄露用户的偏好信息。此外，还可以考虑与其他隐私保护技术相结合，如同态加密等，从而在保证高效运行的同时提升结果的准确性，实现性能与精度的双重增强。

参考文献

[1] Candès, E. and Recht, B. (2012) Exact Matrix Completion via Convex Optimization. *Communications of the ACM*, **55**, 111-119. https://doi.org/10.1145/2184319.2184343

[2] Ji, H., Liu, C., Shen, Z. and Xu, Y. (2010) Robust Video Denoising Using Low Rank Matrix Completion. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, 13-18 June 2010, 1791-1798. https://doi.org/10.1109/cvpr.2010.5539849

[3] Polania, L.F., Carrillo, R.E., Blanco-Velasco, M. and Barner, K.E. (2011) Matrix Completion Based ECG Compression. 2011 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, 30 August-3 September 2011, 1757-1760. https://doi.org/10.1109/iembs.2011.6090502

[4] Gu, S., Zhang, L., Zuo, W. and Feng, X. (2014) Weighted Nuclear Norm Minimization with Application to Image Denoising. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 2862-2869. https://doi.org/10.1109/cvpr.2014.366

[5] Lohr, S. (2010) Netflix Cancels Contest after Concerns Are Raised about Privacy. New York Times.

[6] Mironov, I. and McSherry, F. (2009) Differentially Private Recommender Systems: Building Privacy into the Netflix Prize Contenders. *Proceedings of the* 15*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD*), Paris, 28 June-1 July 2009,627-636.

[7] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I. and Naor, M. (2006) Our Data, Ourselves: Privacy via Distributed Noise Generation. In: Vaudenay, S., Ed., *Advances in Cryptology—EUROCRYPT* 2006, Springer, 486-503. https://doi.org/10.1007/11761679_29

[8] Wang, Z., So, H.C. and Liu, Z. (2022) Fast and Robust Rank-One Matrix Completion via Maximum Correntropy Criterion and Half-Quadratic Optimization. *Signal Processing*, **198**, Article ID: 108580.

https://doi.org/10.1016/j.sigpro.2022.108580

[9] Kang, Y.Y. (2016) Robust and Scalable Matrix Completion. 2016 *International Conference on Big Data and Smart Computing* (*BigComp*), Hong Kong, 18-20 January 2016, 46-52. https://doi.org/10.1109/bigcomp.2016.7425800

[10] Ma, R., Barzigar, N., Roozgard, A. and Cheng, S. (2014) Decomposition Approach for Low-Rank Matrix Completion and Its Applications. *IEEE Transactions on Signal Processing*, **62**, 1671-1683. https://doi.org/10.1109/tsp.2014.2301139

[11] Tzagkarakis, C., Becker, S. and Mouchtaris, A. (2014) Joint Low-Rank Representation and Matrix Completion under a Singular Value Thresholding Framework. 2014 22*nd European Signal Processing Conference* (*EUSIPCO*), Lisbon, 1-5 September 2014, 1202-1206.

[12] Dorffer, C., Puigt, M., Delmaire, G. and Roussel, G. (2017) Fast Nonnegative Matrix Factorization and Completion Using Nesterov Iterations. In: Tichavský, P., Babaie-Zadeh, M., Michel, O. and Thirion-Moreau, N., Eds., *Latent Variable Analysis and Signal Separation*, Springer, 26-35. https://doi.org/10.1007/978-3-319-53547-0_3

[13] Cabral, R., De la Torre, F., Costeira, J.P. and Bernardino, A. (2013) Unifying Nuclear Norm and Bilinear Factorization Approaches for Low-Rank Matrix Decomposition. 2013 *IEEE International Conference on Computer Vision*, Sydney, 1-8 December 2013, 2488-2495. https://doi.org/10.1109/iccv.2013.309

[14] Candès, E.J. and Recht, B. (2009) Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, **9**, 717-772. https://doi.org/10.1007/s10208-009-9045-5

[15] Nie, F., Wang, H., Huang, H. and Ding, C. (2013) Joint Schatten p-Norm and ℓ_p-Norm Robust Matrix Completion for Missing Value Recovery. *Knowledge and Information Systems*, **42**, 525-544. https://doi.org/10.1007/s10115-013-0713-z

[16] Liu, Y., Jiao, L.C. and Shang, F. (2013) A Fast Tri-Factorization Method for Low-Rank Matrix Recovery and Completion. *Pattern Recognition*, **46**, 163-173. https://doi.org/10.1016/j.patcog.2012.07.003

[17] Halko, N., Martinsson, P.G. and Tropp, J.A. (2011) Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Review*, **53**, 217-288. https://doi.org/10.1137/090771806