Advances in Applied Mathematics 应用数学进展, 2026, 15(11), 236-247
Published Online November 2026 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2026.1511446

Hans汉斯

Krylov低秩子空间的随机草图算法

旷　荣

成都理工大学数学科学学院，四川　成都

收稿日期：2026年10月19日；录用日期：2026年11月13日；发布日期：2026年11月19日

摘　要

大规模线性反问题广泛应用于医学成像、图像去模糊以及遥感探测等领域。近年来，核范数正则化方法(NNR)近年来被广泛应用于低秩矩阵恢复及大规模线性逆问题求解中。然而，对于高维问题，由于系统矩阵规模巨大以及核范数正则化所引入的迭代权更新过程，传统迭代重加权核范数方法仍面临较高的计算成本。为有效求解此类问题，已有研究结合Krylov子空间迭代方法提出了IRN-GMRES-NNRp方法，通过在每次外层迭代中构造加权线性系统，并利用GMRES方法求解相应子问题，从而避免直接求解大规模正则化系统。然而，在大规模场景下，IRN-GMRES-NNRp中Arnoldi过程涉及的大量Krylov基向量存储、正交化以及残差计算仍然成为限制算法效率的主要因素。针对上述问题，本文提出一种基于随机草图技术的加IRN-GMRES-NNR方法(SINR-GMRES-NNRp)。通过引入随机草图矩阵对Krylov子空间进行压缩，降低因为正交化和基向量存储所带来的负担从而有效地提升算法效率。数值实验表明，与原始IRN-GMRES-NNR方法相比，所提出方法能够有效降低计算时间和存储需求，同时保持相近的恢复精度和收敛性能，验证了随机草图技术在大规模低秩矩阵恢复问题中的有效性。

关键词

Krylov子空间，随机草图，IRN

Random Sketch Algorithm for Krylov Low-Rank Subspaces

Rong Kuang

School of Mathematical Sciences, Chengdu University of Technology, Chengdu Sichuan

Received: October 19, 2026; accepted: November 13, 2026; published: November 19, 2026

Abstract

Large-scale linear inverse problems are widely applied in fields such as medical imaging, image deblurring, and remote sensing. In recent years, nuclear norm regularization (NNR) methods have

文章引用: 旷荣. Krylov 低秩子空间的随机草图算法[J]. 应用数学进展, 2026, 15(11): 236-247.
DOI: 10.12677/aam.2026.1511446

been extensively used for low-rank matrix recovery and the solution of large-scale linear inverse problems. However, for high-dimensional problems, due to the enormous size of the system matrix and the iterative weight-update process introduced by nuclear norm regularization, traditional iterative reweighted nuclear norm methods still face high computational costs. To effectively solve such problems, previous research has combined Krylov subspace iterative methods to propose the IRN-GMRES-NNRp method, which constructs a weighted linear system at each outer iteration and uses the GMRES method to solve the corresponding subproblems, thereby avoiding the direct solution of large-scale regularized systems. Nevertheless, in large-scale scenarios, the Arnoldi process in IRN-GMRES-NNRp involves substantial storage, orthogonalization, and residual computation of Krylov basis vectors, which remain the main factors limiting algorithm efficiency. To address these issues, this paper proposes a random sketching-based IRN-GMRES-NNR method (SINR-GMRES-NNRp). By introducing a random sketching matrix to compress the Krylov subspace, the burden caused by orthogonalization and basis vector storage is reduced, effectively improving algorithmic efficiency. Numerical experiments show that, compared with the original IRN-GMRES-NNR method, the proposed method can significantly reduce computation time and storage requirements while maintaining similar recovery accuracy and convergence performance, verifying the effectiveness of random sketching technology in large-scale low-rank matrix recovery problems.

Keywords

Krylov Subspace, Random Sketch, IRN

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

本文主要考虑如下的大规模线性反问题：

$$Ax = b, \text{其中 } A \in R^{m\times n}, x \in R^n, b = b^{ex} + n, b \in R^m \tag{1}$$

这里，矩阵 A 表示离散化后的线性前向算子，x 为待恢复的未知变量，b^{ex} 为理想观测数据，而 n 表示测量过程中不可避免引入的噪声。未知向量 $x \in R^n$ 是通过将未知真实图像 $X \in R^{n_1\times n_1}$ 的列重新堆叠而获得的，其中 $n_1 = \sqrt{n}$，我们分别用 $x = \mathrm{vec}(X)$ 和 $X = \mathrm{vec}^{-1}(x)$ 来表示该运算以及其逆。本文主要研究当 A 为方阵时问题(1)的低秩正则化展开[1]。

大量研究表明，自然图像经过适当组织后通常具有明显的低秩特性，其主要信息集中于少数主导奇异值所对应的低维子空间。因此，在反问题求解过程中引入低秩约束，不仅能够抑制噪声干扰，而且能够有效保持图像整体结构，提高恢复精度，直接最小化矩阵秩函数虽然能够获得最优的低秩表示，但由于秩函数具有非凸、非连续等特点，相应优化问题通常属于 NP-hard 问题，难以直接求解[2]。为此，研究者提出利用核范数(nuclear norm)作为矩阵秩函数的凸替代，从而将低秩恢复转化为可有效求解的凸优化问题[3] [4]。对于模型(1)，相应的核范数正则化模型可表示为[1]

$$\min_x \left\|Ax - b\right\|_2^2 + \lambda \left\|\mathrm{vec}^{-1}(x)\right\|_* \tag{2}$$

其中 $\lambda > 0$ 是正则化参数，$\|\cdot\|_*$ 表示矩阵核范数。围绕核范数模型，研究者提出了大量高效的数值算法。Recht、Fazel 和 Parrilo 首次系统建立了核范数最小化的理论基础，证明了核范数是矩阵秩函数在算子范数单位球上的最紧凸包，为低秩恢复问题提供了重要理论依据[3]。Candès 与 Recht 进一步证明，在满足

一定采样条件时仅通过核范数最小化即可精确恢复低秩矩阵[4]。在算法方面，Cai 等提出了奇异值阈值算法(Singular Value Thresholding, SVT) [5]，利用核范数的近端算子实现奇异值的软阈值收缩。尽管核范数模型具有良好的理论性质，但其对所有奇异值采用相同的惩罚策略，往往会导致较大的奇异值被过度收缩，从而影响重建结果的保真度。为了克服这一不足，越来越多的研究开始采用非凸低秩正则化模型，以获得更加接近矩阵真实秩函数的惩罚形式[6]-[9]。

其中，Schatten-p 范数($0 < p \le 1$)由于能够对不同奇异值施加非均匀惩罚，被认为比核范数更加接近矩阵秩函数。基于这一思想，文献[1]提出 Nuclear Norm Regularization-p (NNRp)模型

$$\min_{x} \|Ax-b\|_2^2 + \lambda \left\|\mathrm{vec}^{-1}(x)\right\|_{*,p}, \|X\|_{*,p} = \sum_{i=1}^{n} \left(\sigma_i(X)\right)^p, 0 < p \le 1 \tag{3}$$

当 $p=1$ 时，模型(3)退化为经典核范数模型(2)；而当 $0 < p < 1$ 时，该模型能够更加准确地逼近矩阵秩函数，因此通常具有更优的低秩恢复能力[1]。针对 Schatten-p 等非凸奇异值正则化模型，Lu 等提出迭代重加权核范数方法(IRNN)，通过构造加权核范数替代子问题并进行加权奇异值阈值化，实现对非凸低秩模型的有效求解[10]；其期刊扩展版进一步完善了多块变量情形及收敛性分析[11]。在此基础上，Wang 等引入 Nesterov 型外推与自适应平滑参数更新，提出加速的迭代重加权核范数方法，从而提升算法效率并实现有限步秩识别[12]。与此同时，ADMM 通过变量分裂将复杂优化问题分解为多个易于求解的子问题，在低秩矩阵恢复、图像修复以及矩阵补全等问题中表现出较好的数值稳定性和算法鲁棒性[13]-[15]。此外，对于 $0 < p < 1$ 的非凸低秩恢复，Mohan 和 Fazel 提出的 IRLS-p 方法依据当前迭代的奇异值构造权重，并通过求解一系列加权 Frobenius 范数最小化子问题逼近平滑 Schatten-p 目标[15]，其能够直接利用当前解矩阵的奇异值动态更新权重矩阵，并具有较好的数值稳定性和理论分析基础。并且近年来针对 IRN 的快速算法、收敛理论以及自适应权重更新策略也取得了进一步发展，为大规模低秩矩阵恢复提供了更加高效的求解框架[16]。由于该类子问题可改写为广义 Tikhonov 正则化问题，Gazzola 等[1]进一步采用预条件 Krylov 子空间方法进行高效求解。对于方阵系统，文献[1]中提到了方法 IRN-GMRES-NNRp：改方法在每次外层 IRN 迭代中固定权重，并在内层采用 GMRES 近似求解相应的重加权子问题；随着迭代推进，权重依据当前近似解的奇异值信息更新，使所得解逐步逼近平滑的 Schatten-p 正则化模型。

尽管 IRN-GMRES-NNRp 能够有效求解 Schatten-p 正则化的低秩恢复问题，但在大规模情形下，反复进行矩阵–向量乘法、Krylov 基向量正交化及权重更新仍会带来较高的计算与存储开销。为降低部分高维计算的成本，本文将近年来受到广泛关注的随机草图技术嵌入 IRN-GMRES-NNRp 的 Krylov 子空间迭代过程中，利用低维草图空间中的近似内积替代相应的高维运算。当草图矩阵满足相关子空间嵌入性质且草图维数足够大时，所得方法能够以高概率保持与原算法相近的残差和重建精度。基于此，本文构造了 SIRN-GMRES-NNRp 算法，旨在降低迭代计算成本并提升整体求解效率。本文其余部分安排如下：第二节主要介绍了 IRN-GMRES-NNRp 算法，第三节介绍了我们如何在已有的 IRN-GMRES-NNRp 方法上做出进一步的改进从而得出我们的 SIRN-GMRES-NNRp 算法，第四节主要是提出的算法进行相应的数值实例的实现。

2. IRN-GMRES-NNRp 算法

近年来，Krylov 子空间方法因其仅需矩阵–向量乘法即可完成迭代计算，不需要显式构造矩阵分解，因此已成为求解大规模线性系统最重要的迭代算法之一。其中，广义最小残差算法(Generalized Minimal Residual, GMRES)由 Saad 和 Schultz 提出[17] [18]，并广泛应用于偏微分方程离散、反问题求解以及科学计算等领域。但是当未知解具有明显低秩结构时，GMRES 方法仍然面临着挑战，主要在于 Arnoldi 正交

化、矩阵–向量乘法等基本运算会导致中间变量的秩不断增加，随着迭代深入，存储需求不仅会增加，也会显著提高矩阵运算成本。针对这一问题，文献[19]将低秩截断策略嵌入 GMRES 迭代过程，提出了随机重启低秩 GMRES (RS-LR-GMRES)算法。RS-LR-GMRES 在生成新的 Krylov 基向量后立即执行低秩投影，使所有中间变量始终保持较低秩结构，因此能够有效缓解迭代过程中秩不断增长的问题。然而，由于经截断后的基向量一般不再保持严格正交，相应生成的搜索空间也不再是经典意义上的 Krylov 子空间。文献[1]进一步将低秩 Krylov 方法推广至核范数正则化反问题，提出了 IRN-GMRES-NNRp 低秩求解器，充分利用 Krylov 子空间方法在大型反问题中的计算优势，同时结合 IRLS 框架实现低秩约束与正则化迭代的协同优化。

2.1. Schattenp 问题的平滑与一阶条件

针对第一节提到的 NNRp 问题(3)我们谈到当 $p=1$ 时问题会转化为形式(2)，而当 $0<p<1$ 时问题为非凸且非光滑，为得到可计算的迭代形式，文献[1]以平滑 Schatten 函数 $C_p^{\gamma}(X)$ 替代原正则项，并由当前近似解 $X_k=U_k\Sigma_k V_k^{\mathrm{T}}$ 的奇异值分解构造迭代相关的变换与权重：

$$C_p^{\gamma}(X)=\mathrm{Tr}\left(\left(X^TX+\gamma I\right)^{\frac{p}{2}}\right),\text{其中 }\gamma>0 \tag{4}$$

其中 $\mathrm{Tr}(\cdot)$ 表示矩阵的秩，$C_p^{\gamma}(X)$ 对于 $p>0$ 时可微 $p\geq 1$ 时是凸函数；从而对于问题(3)则可平滑近似为

$$\min_{X\in R^{n_1\times n_1}}\left\|A(X)-B\right\|_F^2+\lambda C_p^{\gamma}(X) \tag{5}$$

此时公式(5)使得目标函数在全域内变得二阶连续可微。接下来算法就能够使用迭代重加权最小二乘法(IRN)来构造每一步的子问题。由于式(5)直接求解存在一定的难度，借助优化领域问题向线性代数领域的跨越，从而有

$$0=2A^*\left(A(X)-B\right)+\lambda p\left(XX^{\mathrm{T}}+\gamma I\right)^{\frac{p}{2}-1}X \tag{6}$$

其中 $A^*(Y)=A^*Y=A^{\mathrm{T}}\mathrm{vec}(Y)$。该式对 X 为非线性问题，矩阵 $\left(XX^{\mathrm{T}}+\gamma I\right)^{\frac{p}{2}-1}X$ 与未知解有关。IRN 的主要想法是：在第 k 次外层迭代中，将这一矩阵由当前近似解 X_k 固定，从而得到一个可求解的二次问题：

$$X_{k+1}=\left\|A(X)-B\right\|_F^2+\frac{\lambda p}{2}\left\|P_kX\right\|_F^2 \tag{7}$$

其中 $P_k=\left(X_kX_k^{\mathrm{T}}+\gamma I\right)^{\frac{p}{4}-\frac{1}{2}}$，$P_kP_k^{\mathrm{T}}=\left(X_kX_k^{\mathrm{T}}+\gamma I\right)^{\frac{p}{2}-1}$。这里在解决核范数(Nuclear Norm)或 l_p 范数正则化问题时，惩罚项是施加在奇异值上的，结合当前矩阵 X_k 的复杂性，当 $X_k=U_{X_k}\Sigma_{X_k}V_{X_k}^{\mathrm{T}}$ 时，根据 Frobenius 范数在任何正交变换下是不改变的，以及现存基于 Krylov 方法的有效策略来解决 IRLS(-p)序列中的每个二次子问题思想[20] [21]，可将 $\left\|P_kX\right\|_F$ 表示成向量 x 的加权二范数并得到二次子问题：

$$x_{k+1}=\arg\min_x\left\|Ax-b\right\|_2^2+\frac{\lambda p}{2}\left\|\left(W_p^{\gamma}\right)_kC_kx\right\|_2^2 \tag{8}$$

其中 $\left(W_p^{\gamma}\right)_k=I\otimes\left(\Sigma_{X_k}^2+\gamma I\right)^{\frac{p}{4}-\frac{1}{2}}$，$C_k=V_{x_k}^T\otimes U_{x_k}^T$，由于 $0<p\leq 1$ 则 $\frac{p}{4}-\frac{1}{2}<0$ 故存在当奇异值 σ_i 越小时权重越大，反之亦然，从而体现在小奇异值方向会被更重地惩罚，大奇异值方向惩罚较弱。C_k 将 x 映射到当前奇异值域，W_k 执行这一自适应重加权。

2.2. IRN-GMRES-NNRp 的 Arnoldi 投影

IRN-GMRES-NNRp 适用于 A 为方阵的情形。在第 k 个外层循环中，C_k 与 W_k 固定。

取 $x'=W_kC_kx$，$x=C_k^{\mathrm{T}}W_k^{-1}x'$，由于 C_k 为正交矩阵，子问题可变为标准形式的 Tikhonov 正则化问题，再对变换后的矩阵 $C_kAC_k^{\mathrm{T}}W_k^{-1}$ 实施 Arnoldi 过程。相应的投影关系和近似解如下：

$$\left(C_kAC_k^{\mathrm{T}}W_k^{-1}\right)V_m=V_{m+1}H_m,x_m=C_k^{\mathrm{T}}W_k^{-1}V_my_m \tag{9}$$

$$y_m=\arg\min_y\left\|H_my-\|b\|_2e_1\right\|_2^2+\lambda_m'\|y\|_2^2 \tag{10}$$

这里的 λ_m' 是投影问题中的正则化参数，可在 hybrid Krylov 过程中选择。由于 C_k 正交，残差范数也能由小规模 Hessenberg 矩阵直接计算：

$$\|b-Ax_m\|_2=\left\|C_kb-\hat{A}_k\hat{x}_m\right\|_2=\left\|\|b\|_2e_1-H_my_m\right\| \tag{11}$$

其中 $\hat{A}_k=C_kAC_k^{\mathrm{T}}W_k^{-1},\hat{x}_m=V_my_m$。因此，IRN-GMRES-NNRp 的低秩机制来自固定点重加权，而 GMRES 的作用是高效近似每一个固定权重二次子问题的解。文献[1]中的“预处理”主要是将低秩先验加入解空间，并非只为加快线性迭代收敛。上述的推导即为 IRN-GMRES-NNRp 的理论创新。

3. SIRN-GMRES-NNRp 算法

随机草图作为随机数值线性代数的核心技术，近年来已经从一门新兴的理论学科发展为极具工业落地价值的研究方向，作为随机代数数值的特殊技巧受到广发的关注[22] [23]，其主要思想是尝试在整个观测系统的“缩影”上去来解决问题，该思想通过子空间嵌入思想即借助一个更低维度的矩阵 $\Omega\in R^{s\times m}$，使得对于所有的 $x\in\Gamma$ 都有下式成立：

$$(1-\varepsilon)\|x\|_2\le\|\Omega x\|_2\le(1+\varepsilon)\|x\|_2 \tag{12}$$

则称 Ω 为子空间 Γ 的很好的草图，其中 s 远远小于 m，ε 是畸变参数，一般情况下 $\varepsilon\in(0,1)$；这就是线性降维技术的主要思想，从而对于草图的选择主要遵循式(15)。将草图融入经典数值算法(如 GMRES、LSQR、Arnoldi 过程)以加速偏微分方程(PDE)和大型物理反演的求解，是近两年的另一大热点。Burke 等人借助该一思想开发出了 GMRES-SDR 算法，利用调和 Ritz 向量的草图化表示实现了极低开销的子空间回收[24]，D Palitta 和 M Pasha 针对大规模 Tikhonov 正则化问题，该文提出了 sGKS 的算法，解决了传统广义 Krylov 子空间(GKS)方法中矩阵的 QR 分解计算量大，以及全空间正交化带来的开销[25]。在核范数正则化问题中，传统的 IRN-GMRES-NNRp 方法通过迭代重加权策略将非线性的核范数约束转化为一系列线性子问题，并利用 Krylov 子空间方法求解每一步线性系统。但是，当矩阵 A 为大规模稀疏矩阵时，Arnoldi 过程中涉及的大量高维向量内积和正交化操作会成为主要计算瓶颈。因此，为降低 Krylov 子空间计算成本，本文引入随机草图技术，对 Krylov 子空间进行低维随机嵌入。设 x_0 为当前内层初值，则有：

$$r_0=b-Ax_0,\beta=\|r_0\|_2,v_1=\frac{r_0}{\beta} \tag{13}$$

在第 j 步定义：

$$D_K=\left(\Sigma_{X_k}^2+\gamma I\right)^{\frac{p}{4}-\frac{1}{2}},P_k'=C_k^{\mathrm{T}}\left(W_p^{\gamma}\right)_k^{-1}$$

$$q_j=P_k'v_j,Q_j=\left[q_1,q_2,\cdots,q_j\right]=P_k'V_j \tag{14}$$

其中设当前所有可能残差属于子空间

$$u_j=\mathrm{span}\left\{r_0,AQ_j\right\}$$

则此时残差为：

$$\begin{aligned} r_j^{\Omega} &= \Omega r_j(y) = \Omega\left(b - C_k A x_j(y)\right) \\ &= \Omega\left(r_0 - C_k A P_k' V_j y\right) \\ &= \Omega\left(V_{j+1}\left(\beta e_1 - H_j y\right)\right) \\ &= g_j - G_j y \end{aligned}$$

其中 $G_j = \Omega V_{j+1} H_j$， $g_j = \Omega r_0$，并且如果随机矩阵 Ω 在 u_j 上近似保持二范数；那么对于任意的 y 都有：

$$(1-\varepsilon)\left\|g_j - G_j y\right\|_2^2 \le \left\|g_j - G_j y\right\|_2^2 \le (1-\varepsilon)\left\|g_j - G_j y\right\|_2^2$$

因此，可以用草图残差最小化近似替代原始残差最小化。由于重加权方向存在 Q_j 中，则有：

$$x_j^{\Omega} = x_0 + Q_j y_j^{\Omega}$$

其中

$$y_j^{\Omega} = \arg\min_{y \in R^m} \left\|\Omega\left(r_0 - C_k A Q_j y\right)\right\|_2^2 + \lambda_{\Omega}\|y\|_2^2$$

为草图正则化子问题，主要流程见下列算法：SIRN-GMRES-NNRp 算法。

定理 1：设 IRN-GMRES-NNRp 满足 $AC_k^{\mathrm{T}}\left(W_p^{\gamma}\right)_k^{-1} V_j = V_{j+1} H_j$，若随机草图 S 矩阵满足 $(1-\varepsilon)\|x\|_2 \le \|\Omega x\|_2 \le (1+\varepsilon)\|x\|_2$ 则 SIRN-GMRES-NNR 得到的残差 r_j^{Ω} 满足：

$$\left\|r_j^{\Omega}\right\|_2 \le \frac{1+\varepsilon}{1-\varepsilon}\left\|r_j^*\right\|_2$$

其中 r_j^* 为 IRN-GMRES-NNRp 残差。

证明：根据 JL 性质有

$$(1-\varepsilon)\left\|r_j^{\Omega}\right\|_2 \le \left\|\Omega r_j^{\Omega}\right\|_2$$

由于 y_j^{Ω} 最小化残差可以得到：

$$\left\|\Omega r_j^{\Omega}\right\|_2 \le \left\|\Omega r_j^*\right\|_2$$

再次利用 JL 可以得到：

$$\left\|\Omega r_j^*\right\|_2 \le (1+\varepsilon)\left\|r_j^*\right\|_2$$

因此：

$$(1-\varepsilon)\left\|r_j^{\Omega}\right\|_2 \le (1+\varepsilon)\left\|r_j^*\right\|_2$$

进一步有：

$$\left\|r_j^{\Omega}\right\|_2 \le \frac{1+\varepsilon}{1-\varepsilon}\left\|r_j^*\right\|_2$$

证毕。

SIRN-GMRES-NNRp 算法：

1. 输入：算子 A，观测数据 b，真值 x_{ex}，初值 x_0，maxit，cycles 等参数；
2. 初始化：$U_x, V_x, S_x = I$；
3. **for** $i = 1$，...cycles 外层：
   1) 更新参数集：$opt = \{U_x, V_x, S_x, \cdots\}$
   2) $r_0 = b - Ax_0, \beta = \|r_0\|_2, v_1 = \dfrac{r_0}{\beta}$
   3) **for** $k = 1$，...cycles 内层：
      a) 计算预处理方向：$Z_k = \text{Precond}(v_k)$，$Z_k = [z_1, z_2, \cdots, z_k]$
      b) 确定搜索方向：$w = Az_k$
      c) 执行草图 Arnoldi 过程 $H_{i,k} = (\Omega V_i)^{\mathrm{T}} (\Omega(AZ_k))$；
      d) 正交化并更新基：$w \leftarrow w - V_k H(1:k,k)$，$\Omega v_{k+1} = \dfrac{\Omega w - \Omega V_k H(1:k,k)}{H(k+1,k)}$
      e) 对草图后的投影算子做 SVD：$[U_H, \Sigma_H, V_H] = \text{svd}(\Omega V_{k+1} H_k)$
      f) 求解草图空间极小二乘问题：$y_m^{\Omega} = \arg\min_{y \in R^m} \|\Omega(r_0 - C_k A Q_m y)\|_2^2 + \lambda_m \|y\|_2^2$
      g) 更新近似解 x_k；
      h) 返回解序列 x_k 及残差；
      i) 终止内循环
   4) 基于当前解 x_i 的矩阵的 SVD 分解更新 V_k, U_X, Σ_X；
   5) 若满足停机准则，跳出循环
4. 输出：最优解及收敛历史

其中 Precond 是 NNRp正则项在当前解处的白化算子：IRN 的核心是将非平滑的惩罚项转化为加权的二范数项，IRN 将平滑 Schatten-p拟范数 $\sigma_{x,i}^{\frac{p-2}{2}}$ 用二次代理 $w_i = \left(\dfrac{\sigma_{x,i}}{\sigma_{x,i}^2 + \delta^2}\right)^{\frac{2-p}{2}} \sigma(X)^2$ 逼近，σ_x 为解矩阵的奇异值，当 $p < 2$ 时小奇异值方向被赋予大权重而被强烈压制，从而实现低秩促进；其中 $\delta = \max(\sigma_x) \cdot 10^{-4}$ 起平滑作用，避免谱截断后或解秩亏时权重奇异。

4. 数值实例

在本节中我们进行了相应的数值实例，通过实验得到的时间和迭代次数与相对历史残差图来比较我们的算法 SIRN-GMRES-NNRp 与 IRN-GMRES-NNRp、FGMRES-NNRp、FLSQR-NNRp、FGMRES-NNRp 等方法的性能，不同的实验我们通过设置不同的最大迭代次数从而使各方法能够在较为一致的实验条件下进行性能比较。

实验 4.1：第一组实验我们使用 IR Tools 工具包中的数据测试了 256 维的相应实验数据，考虑当 $p = 0.75$ 时，对于 IRN-GMRES-NNRp 和 SIRN-GMRES-NNRp 而言，内层最大迭代次数设置为 200，当内循环满足差异原则时即可启动新的外循环，其中外循环一共有 5 次。该实验中涉及的所有方法我们都采取纯迭代版本，即使得正则化参数为 0，除以上阐述外其余相应的参数设置我们保留和原始的代码一致。在图 1 中我们展示了重建后的效果图像，其中包括精确与模糊图形以及 IRN-GMRES-NNRp 和 SIRN-

GMRES-NNRp 的最佳重建效果图，图 2 中展示了对于所有实验的相对残差历史图；图 3 中展示了在满足上述条件的情况下其相对残差历史随着时间的变化趋势图，并且为了更好地观察实验后我们的算法效果，通过设置表格将我们的算法与原本的算法运行结束后的时间进行对比。图 2 与图 3 分别给出了 $n = 65536$ 时各比较方法的相对残差随迭代次数和随计算时间的变化，表 1 进一步汇总了各方法的总运行时间与最终相对残差。

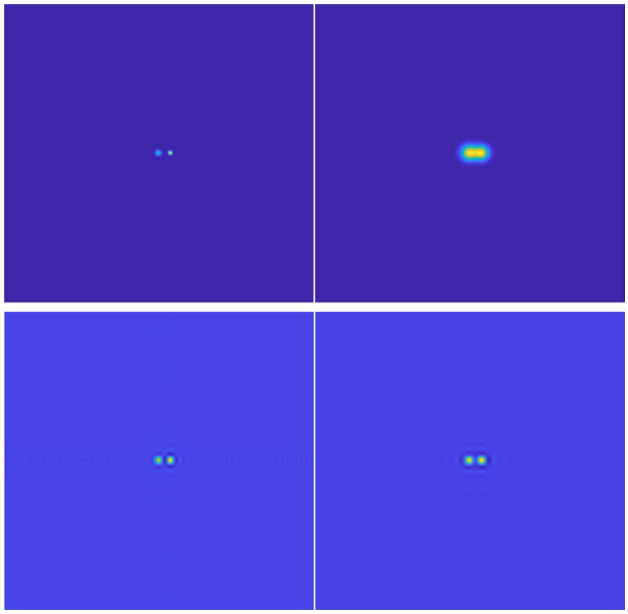


Figure 1. 256 × 256-dimensional reconstructed images: the first row from left to right shows the ground truth image and the noisy image; the second row from left to right shows the best reconstructed image by INR-GMRES-NNR and the best reconstructed image by SINR-GMRES-NNR
图 1. 256 × 256 维重建图像，第一排从左到右依次为精确图、噪声图；第二排从左到右依次为 INR-GMRES-NNR 最佳重建图与 SINR-GMRES-NNR 最佳重建图

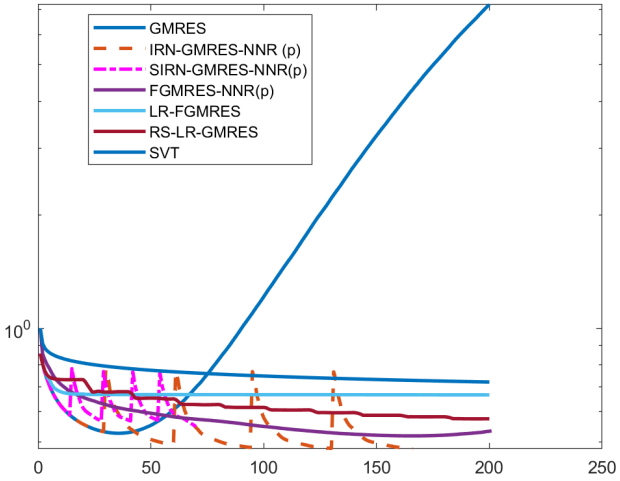


Figure 2. Plot of relative residual history versus number of iterations when n = 65536
图 2. n = 65536 时相对残差随迭代次数的变化图

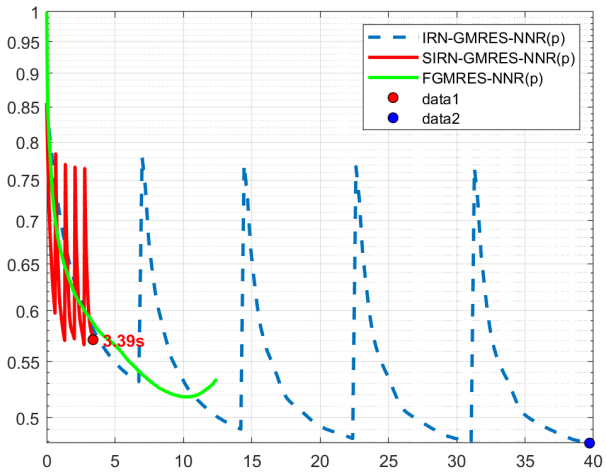


Figure 3. Graph of the relative residual changes over time when n = 65536
图 3. n = 65536 时相对残差随时间的变化图

Table 1. Algorithm performance comparison summary when n = 65536
表 1. n = 65536 时算法性能对比总结

方法	时间(s)	残差
IRN-GMRES-NNRp	39.7222	4.7870e−01
SINR-GMRES-NNRp	3.3901	5.7115e−01
FGMRES-NNRp	12.3676	5.3292e−01
IRN-GMRES-NNRp 与 SINR-GMRES-NNRp 的加速比：11.72x		

由图 1 可见，本文提出的 SIRN-GMRES-NNRp 在整个迭代过程中保持了与其对应原始方法几乎一致的下降趋势与最终残差水平，随机化 sketching 的引入并未带来明显的振荡或精度退化；相比之下，SVT 在迭代后期出现了相对残差的急剧回升，体现出纯迭代正则化典型的半收敛现象。图 2 进一步表明，在残差水平相当的条件下，随机化方法所需的计算时间显著更少，其中以内层求解占主导的 IRN 类方法收益最为突出。定量地，表 1 显示 SIRN-GMRES-NNRp 相对 IRN-GMRES-NNRp 取得了 11.72 倍的加速。上述结果说明，引入随机 sketching 能够有效降低基正交化与残差计算的开销，同时基本保持原始方法的收敛行为与重建质量，这一优势在大规模图像反问题中尤为显著。

实验 4.2：第二组实验我们使用了 IR Tools 工具包中的数据，进行 512 维图像的测试，在该实验中我们将内层最大迭代次数设置为 600，外循环一共设置了有 10 次，其他参数设置和实验 4.1 保持一致。同样在图 4 中我们展示了当前重建后的效果图像，其中包括精确与模糊图形以及 IRN-GMRES-NNRp 和 SIRN-GMRES-NNRp 的最佳重建效果图，图 5 与图 6 分别给出了 n = 262,144 (即 512 测试图像)时各比较方法的相对残差随迭代次数和随计算时间的变化，表 2 进一步汇总了各方法的总运行时间、最终相对误差及相应的加速比。由图 5 可见，本文提出的方法在整个迭代过程中保持了与其对应原始方法几乎一致的收敛轨迹与最终残差水平，随机化 sketching 的引入并未造成明显的振荡或精度退化；与之形成鲜明对比的是，SVT 等若干对比方法在迭代后期出现了相对残差的显著回升，呈现出纯迭代正则化典型的半收敛现象。进一步表明，在残差水平相当的条件下，随机化方法只需远少于原始方法的计算时间即可达到

同一收敛阶段，图中标注的终止时刻亦直观地体现了这一差距。定量地，表 2 显示 SIRN-GMRES-NNRp 相对 IRN-GMRES-NNRp 取得了 47.83 倍的。值得注意的是，与 $n = 65{,}536$ 时的结果相比，随机化方法的加速比在更大规模问题上均显著提升，说明随机 sketching 所带来的收益随问题规模增长而放大。

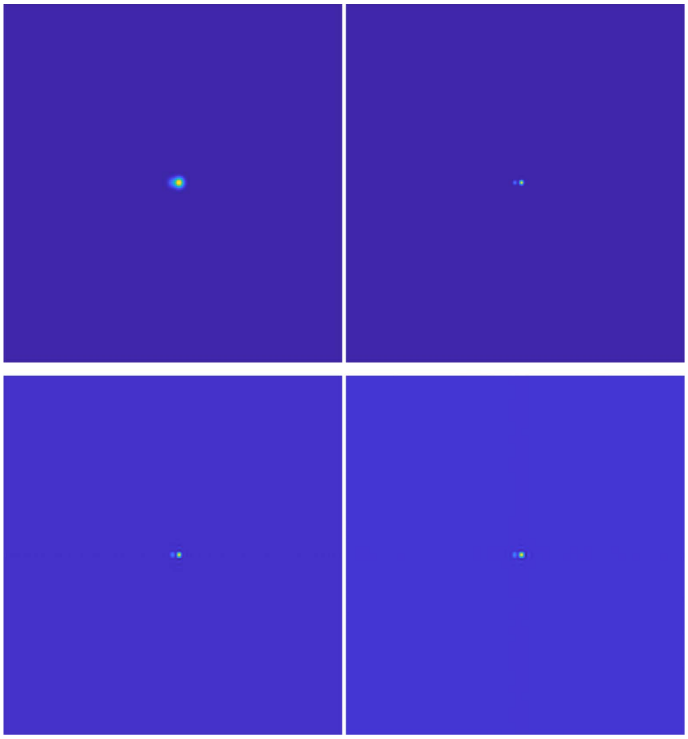


Figure 4. Reconstructed images: the first row from left to right shows the ground truth image and the noisy image; the second row from left to right shows the best reconstructed image by INR-GMRES-NNR and the best reconstructed image by SINR-GMRES-NNR
图 4. 重建图像，第一排从左到右依次为精确图、噪声图；第二排从左到右依次为 INR-GMRES-NNR 最佳重建图与 SINR-GMRES-NNR 最佳重建图

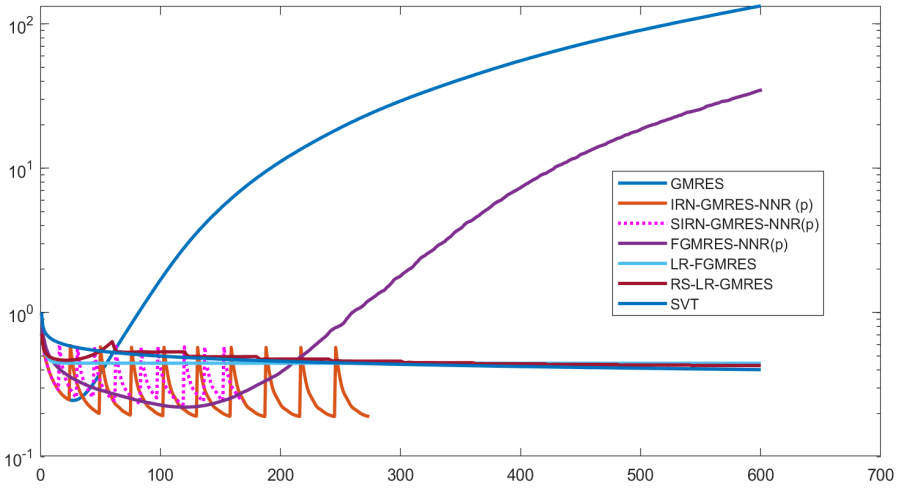


Figure 5. Plot of relative residual history versus number of iterations
图 5. 相对残差随迭代次数的变化图

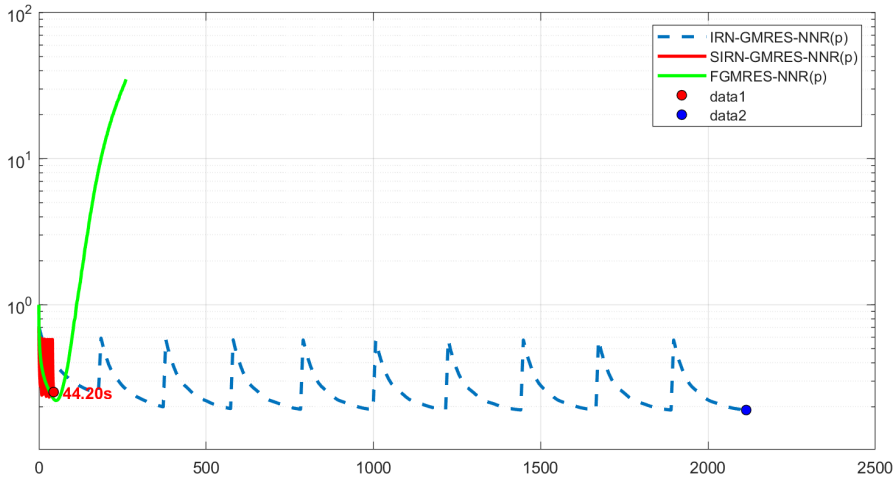


Figure 6. Graph of the relative residual changes over time
图 6. 相对残差随时间的变化图

Table 2. Algorithm performance comparison summary
表 2. 算法性能对比总结

方法	时间(s)	残差
IRN-GMRES-NNRp	2113.9436	1.9017e−01
SINR-GMRES-NNRp	44.1960	2.5241e−01
FGMRES-NNRp	260.9388	3.3931e+01
IRN-GMRES-NNRp 与 SINR-GMRES-NNRp 的加速比：47.83x		

上述结果充分表明，本文方法能够在基本保持原始算法收敛行为与重建质量的前提下大幅降低基正交化与残差计算的开销，因而尤其适用于大规模图像反问题。

5. 结论

本文我们主要针对 IRN-GMRES-NNRp 和 FGMRES-NNRp 方法进行了改进，主要针对当前向算子十分大时，我们将现有流行的草图压缩思想也结合运用到了我们 Krylov 子空间中，保证在相对残差基本相当的情况下来加快我们的运行速度，对此我们进行了两组实验分别为 $n = 65536$ 和 $n = 262144$，并将我们的算法和先前的一些算法进行了结果对比，对此我们得到两组实验在运行速度上在保证相对历史残差基本和原本的算法相对历史残差相当的情况下均能显著提高其运行的速度。

参考文献

[1] Gazzola, S., Meng, C. and Nagy, J.G. (2020) Krylov Methods for Low-Rank Regularization. *SIAM Journal on Matrix Analysis and Applications*, **41**, 1477-1504. https://doi.org/10.1137/19m1302727

[2] Fazel, M., Hindi, H. and Boyd, S.P. (2001). A Rank Minimization Heuristic with Application to Minimum Order System Approximation. *Proceedings of the* 2001 *American Control Conference* (*Cat. No.*01CH37148), 25-27 June 2001, Arlington, 4734-4739. https://doi.org/10.1109/acc.2001.945730

[3] Recht, B., Fazel, M. and Parrilo, P.A. (2010) Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review*, **52**, 471-501. https://doi.org/10.1137/070697835

[4] 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

[5] Cai, J., Candès, E.J. and Shen, Z. (2010) A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization*, **20**, 1956-1982. https://doi.org/10.1137/080738970

[6] Fornasier, M., Rauhut, H. and Ward, R. (2011) Low-Rank Matrix Recovery via Iteratively Reweighted Least Squares Minimization. *SIAM Journal on Optimization*, **21**, 1614-1640. https://doi.org/10.1137/100811404

[7] Mohan, K. and Fazel, M. (2012) Iterative Reweighted Algorithms for Matrix Rank Minimization. *The Journal of Machine Learning Research*, **13**, 3441-3473.

[8] 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

[9] Lu, C.Y., Lin, Z.C. and Yan, S.C. (2015) Smoothed Low Rank and Sparse Matrix Recovery by Iteratively Reweighted Least Squares Minimization. *IEEE Transactions on Image Processing*, **24**, 646-654. https://doi.org/10.1109/tip.2014.2380155

[10] Lu, C., Tang, J., Yan, S. and Lin, Z. (2014). Generalized Nonconvex Nonsmooth Low-Rank Minimization. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 4130-4137. https://doi.org/10.1109/cvpr.2014.526

[11] Lu, C., Tang, J., Yan, S. and Lin, Z. (2016) Nonconvex Nonsmooth Low Rank Minimization via Iteratively Reweighted Nuclear Norm. *IEEE Transactions on Image Processing*, **25**, 829-839. https://doi.org/10.1109/tip.2015.2511584

[12] Wang, H., Wang, Y. and Yang, X. (2024) Efficient Active Manifold Identification via Accelerated Iteratively Reweighted Nuclear Norm Minimization. *Journal of Machine Learning Research*, **25**, 1-44.

[13] Wang, Y., Zhang, Y. and Fu, X. (2023) Low-Rank Matrix Recovery via Novel Double Nonconvex Nonsmooth Rank Minimization with ADMM. *Multimedia Tools and Applications*, **83**, 15547-15564. https://doi.org/10.1007/s11042-023-16098-y

[14] Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2011) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning*, **3**, 1-122. https://doi.org/10.1561/2200000016

[15] Mohan, K. and Fazel, M. (2010) Iterative Reweighted Least Squares for Matrix Rank Minimization. 2010 48*th Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*), Monticello, 29 September-1 October 2020, 653-661. https://doi.org/10.1109/allerton.2010.5706969

[16] Ge, Z., Zhang, X. and Wu, Z. (2022) A Fast Proximal Iteratively Reweighted Nuclear Norm Algorithm for Nonconvex Low-Rank Matrix Minimization Problems. *Applied Numerical Mathematics*, **179**, 66-86. https://doi.org/10.1016/j.apnum.2022.04.008

[17] Saad, Y. (2003) Iterative Methods for Sparse Linear Systems. Society for Industrial and Applied Mathematics.

[18] Saad, Y. and Schultz, M.H. (1986) GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. *SIAM Journal on Scientific and Statistical Computing*, **7**, 856-869. https://doi.org/10.1137/0907058

[19] Lee, K. and Elman, H.C. (2017) A Preconditioned Low-Rank Projection Method with a Rank-Reduction Scheme for Stochastic Partial Differential Equations. *SIAM Journal on Scientific Computing*, **39**, S828-S850. https://doi.org/10.1137/16m1075582

[20] Renaut, R.A., Vatankhah, S. and Ardestani, V.E. (2017) Hybrid and Iteratively Reweighted Regularization by Unbiased Predictive Risk and Weighted GCV for Projected Systems. *SIAM Journal on Scientific Computing*, **39**, B221-B243. https://doi.org/10.1137/15m1037925

[21] Rodríguez, P. and Wohlberg, B. (2008) An Efficient Algorithm for Sparse Representations with ℓ^p Data Fidelity Term. *Proceedings of* 4*th IEEE Andean Technical Conference* (*ANDESCON*), Cusco, October 2008.

[22] Burke, L., Carson, E. and Ma, Y. (2025) On the Numerical Stability of Sketched GMRES. arXiv: 2503.19086.

[23] Landman, M.S. and Nakatsukasa, Y. (2025) Randomized Flexible Krylov Methods for ℓ_p Regularization. arXiv: 2510.11237.

[24] Burke, L., Güttel, S. and Soodhalter, K.M. (2025) GMRES with Randomized Sketching and Deflated Restarting. *SIAM Journal on Matrix Analysis and Applications*, **46**, 702-725. https://doi.org/10.1137/23m1619472

[25] Palitta, D. and Pasha, M. (2026) A Sketched Generalized Krylov Subspace Method for Large-Scale Regularization. arXiv: 2606.18073.