arXiv:2603.01837v1 [cs.LG] 2 Mar 2026

约束粒子搜索：仅通过前向传播求解扩散逆问题

Hongkun Dou[1*]、**Zike Chen**[1*]、**Zeyu Li**[1],**Hongjue Li**[1†]、**Lijun Yang**[1],**Yue Deng**[1] [2]
[1]北京航空航天大学 [2]北京中关村学院{douhk, zy2415215,
lizeyu123478, lihongjue, yanglijun, ydeng}@buaa.edu.cn。

摘要

扩散模型凭借建模复杂数据分布的能力，已成为求解逆问题的强大生成工具。然而，现有方法通常依赖对前向观测过程的完整了解，以计算用于引导采样的梯度，因此在无法获取此类信息的场景中适用性有限。本文提出约束粒子搜索（***CPS***），这是一种新颖的无梯度方法，它利用所有候选粒子的信息进行主动搜索，以寻找最优粒子，同时纳入与无条件先验高密度区域相一致的约束。不同于以往被动选择有潜力候选项的方法，CPS 将逆问题重新表述为一个约束优化任务，从而实现更灵活、高效的粒子搜索。我们证明，CPS 能够有效求解图像和科学逆问题，取得与基于梯度的方法相当的结果，同时显著优于无梯度替代方法。代码见 https://github.com/deng-ai-lab/CPS。

引言

逆问题是诸多领域面临的基础性挑战，包括计算成像（Tonolini 等人 2020；McCann、Jin 和 Unser 2017）、医学（Song 等人 2021）、地球物理学（Schwarzbach、B¨orner 和 Spitzer 2011）以及流体动力学（Li 等人 2020）。其目标是从间接且含噪的观测 $\boldsymbol{y} \in \mathbb{R}^m$ 中恢复原始信号 $\boldsymbol{x} \in \mathbb{R}^d$，通常可将其建模为：

$$\boldsymbol{y} = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{\eta} \tag{1}$$

其中，$\mathcal{H}(\cdot): \mathbb{R}^d \to \mathbb{R}^m$ 表示前向测量算子，$\boldsymbol{\eta}$ 表示加性噪声。这个问题本质上是不适定的（$m < d$），也就是说，可能有多个解都满足给定的观测。为解决这一问题，适当的先验和正则化技术至关重要，有助于缓解不适定性并确保得到有意义的解。

生成建模领域近期取得的进展，尤其是扩散模型（Ho、Jain 和 Abbeel，2020；Song 等人，2020；Lipman 等人，2022；Rombach 等人，2022），已证明可作为即插即用先验，有效解决逆问题，相较于传统的人工先验（Tibshirani，1996年；Iordache、Bioucas-Dias 和 Plaza，2012年）具有显著优势。现有方法依赖观测过程的梯度或伪逆信息来引导无条件采样过程，确保生成结果与观测相符（Zhang 等人 2025年；Wang、Yu 和 Zhang，2022年；Chung 等人 2022；Kawar 等人，2022a；Mardani 等，2023年）。然而，对这类特权信息的依赖限制了这些方法的适用性，尤其是在现实场景中。例如，在某些场景下，前向模型往往涉及高度非线性或计算成本高昂的数值模拟，因此梯度计算颇具挑战性（Oliver、Reynolds 和 Liu，2008 年；Gillhofer 等人，2019；Duong 等人，2020年；Kawar 等人，2022年b）。

已有多项研究探索利用预训练扩散模型，在只能以黑盒方式访问前向模型的情况下求解逆问题。其中一种方法是 SCG（Huang 等人，2024），它采用了类似拒绝采样的策略。在反向去噪过程的每一步中，SCG 都会采样多个候选粒子，并仅保留与观测最吻合的一个，丢弃其余粒子。尽管这种方法简单直观，但由于每一步都会丢弃许多候选粒子，因此效率低下。另一种值得关注的方法是 EnKG（Zheng 等人，2025a），其灵感来自集合卡尔曼滤波（Katzfuss、Stroud和Wikle 2016）。EnKG 初始化并维护一组粒子来表示状态，并根据观测逐步对其进行优化。然而，在高维逆问题中，EnKG 需要维护数千个粒子，导致巨大的计算开销。此外，DPG（Tang 等人，2024年）通过使用策略梯度避免计算导数，但它需要向干净样本添加噪声扰动，这可能会使样本偏离数据流形。更重要的是，在梯度可用时，与基于梯度的方法相比，这些方法的表现往往较差。

为克服这些局限，我们提出约束粒子寻优（***CPS***），这是一种有原则的方法，综合所有候选粒子中的信息，以识别最优解。我们认为，每个候选粒子都能提供有价值的局部信息，而不应被直接舍弃；这些信息

*这些作者贡献相同。

†通讯作者。版权所有 © 2026，美国人工智能促进协会（www.aaai.org）。保留所有权利。

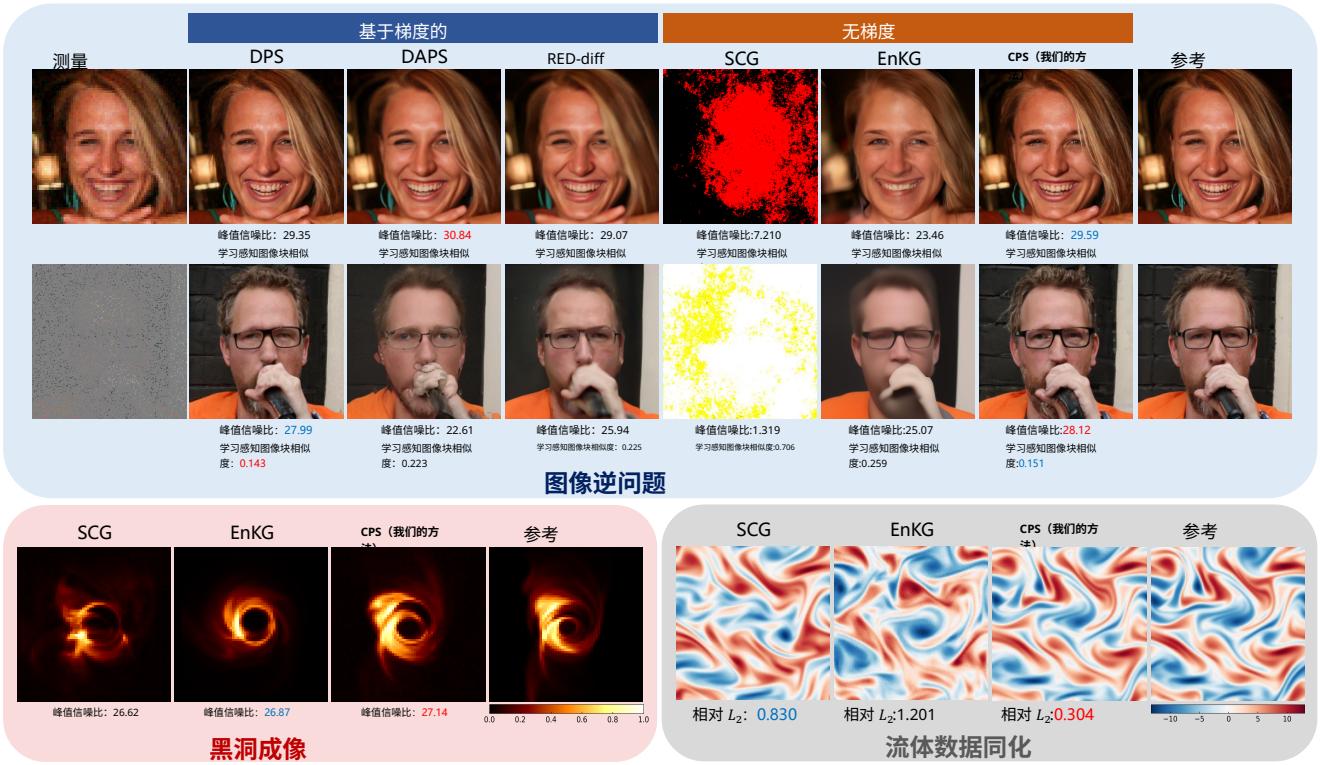


图 1：CPS 与基线方法的定性比较。所提出的 CPS 仅依赖观测模型的前向传播来解决图像和科学逆问题。其表现可与先进的基于梯度的方法相媲美，并显著优于无梯度方法。最佳结果以红色标出，次佳结果以蓝色显示。

能够共同引导去噪过程。在 CPS 中，我们将扩散逆问题重新表述为一个约束优化任务。目标是找到一个与观测最优匹配的粒子，其中匹配程度通过前向测量的局部代理模型来衡量。该代理模型利用全部候选粒子进行稳健估计。同时，我们引入一个约束，确保得到的粒子始终处于逆扩散过程的高概率密度区域内，从而保持与扩散先验的一致性。通过利用所有可用的候选粒子，CPS 提高了采样效率，并在观测保真度与先验知识之间实现了更优的平衡。此外，我们还引入重启策略（Xu 等人，2023；Lugmayr 等人，2023），以迭代修正序贯采样过程中固有的累积误差，从而显著提升最终解的鲁棒性与准确性。

我们在多种任务上评估了 CPS，包括基准图像逆问题和两个科学逆问题：黑洞成像与流体数据同化，这两者均具有高度非线性的测量过程。结果表明，CPS 不仅在图像逆问题上的表现可与常用的基于梯度的方法媲美，还能显著提升无法获得梯度的正向过程中的效率。

预备知识与任务构建

扩散模型

扩散模型（Ho、Jain 和 Abbeel 2020年；Song 等人 2020年）定义了一个随机过程 $\{\boldsymbol{x}_t\}_{t\in[0,T]}$，该过程逐步将数据 $\boldsymbol{x}_0 \sim p_{\text{data}}$ 转换为噪声。该过程由以下微分方程描述：

$$\mathrm{d}\boldsymbol{x}_t = \mathbf{f}(\boldsymbol{x}_t, t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w} \tag{2}$$

其中，$\mathbf{f}(\cdot,\cdot)$ 是漂移系数，$g(\cdot)$ 是扩散系数，$\mathbf{w}$ 表示标准维纳过程。随着时间推移，数据会逐渐扰动为高斯密度 $\boldsymbol{x}_T \sim \mathcal{N}(0, \sigma_T{}^2 I)$。为了从噪声生成数据，扩散模型通过去噪来逆转这一过程，如下所述：

$$\mathrm{d}\boldsymbol{x}_t = \left[\mathbf{f}(\boldsymbol{x}_t, t) - g(t)^2 \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t)\right]\,\mathrm{d}t + g(t)\,\mathrm{d}\bar{\mathbf{w}} \tag{3}$$

其中，$\bar{\mathbf{w}}$ 表示反向时间维纳过程，而 $\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t)$ 是得分函数，它表示时刻 t 边缘分布的对数似然梯度。可以通过最小化去噪得分匹配损失（Vincent，2011）来使用神经网络近似得分函数。

扩散后验采样（DPS）

人们提出了多种方法，利用扩散模型作为即插即用先验来解决逆问题。

一种值得注意的方法是扩散后验采样（DPS）（Chung 等人 2022）。具体而言，DPS 将公式 3 中的无条件得分函数替换为条件得分函数：

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{y}) = \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t) \quad (4)$$

此修改旨在根据观测，从后验分布 $p(\boldsymbol{x}_0|\boldsymbol{y})$ 中进行采样。然而，条件得分函数涉及以下难以计算的积分：

$$p(\boldsymbol{y}|\boldsymbol{x}_t) = \int p(\boldsymbol{y}|\boldsymbol{x}_0)p(\boldsymbol{x}_0|\boldsymbol{x}_t)\,\mathrm{d}\boldsymbol{x}_0 = \mathbb{E}_{p(\boldsymbol{x}_0|\boldsymbol{x}_t)}[p(\boldsymbol{y}|\boldsymbol{x}_0)] \quad (5)$$

该积分计算开销大，通常需要对完整轨迹进行多次采样，以获得干净样本。为解决这一问题，DPS 将引导项近似如下：

$$p(\boldsymbol{y}|\boldsymbol{x}_t) \approx p(\boldsymbol{y}|\hat{\boldsymbol{x}}_{0|t}) = -\frac{1}{\sigma_y^2}\nabla_{\boldsymbol{x}_t}\|\boldsymbol{y} - \mathcal{H}(\hat{\boldsymbol{x}}_{0|t})\|^2 \quad (6)$$

其中，$\hat{\boldsymbol{x}}_{0|t} = \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]$ 可通过 Tweedie 公式结合单步去噪计算得到（Efron，2011年），而 σ_y^2 是观测中噪声的方差。借助这一近似，DPS 无需训练或微调即可解决逆问题。然而，DPS 需要观测算子具有明确的梯度，这限制了它在梯度难以计算或无法获得的场景中的适用性。

将逆问题视为随机控制

我们采用 SCG（Huang 等人，2024）提出的任务表述，并在随机控制框架内构建逆问题的求解方法。为了引导去噪轨迹朝满足观测的目标状态前进，我们在公式 3 的每个时间点 t 引入控制信号 $\boldsymbol{u}_t$。这样便将逆问题转化为随机最优控制问题（Pavon，1989年），其终端代价和瞬时代价表示如下：

$$\mathcal{C} = \mathbb{E}\left[\Phi(\boldsymbol{x}_0) + \int_0^T \frac{\alpha}{2}\|\boldsymbol{u}_t\|^2|\boldsymbol{x}_T\right] \quad (7)$$

其中，$\Phi(\boldsymbol{x}_0)$ 表示终端代价，而 α 是与瞬时代价相关的系数。在逆问题的背景下，我们将其定义为 $\Phi(\boldsymbol{x}_0) = \|\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}_0)\|^2$。通过应用路径积分控制理论（Theodorou、Buchli 和 Schaal，2010年；Williams、Aldrich 和 Theodorou，2017年；Zhang 和 Chen，2023年），我们得到如下最优策略：

$$\mathbf{u}_t^* \mathrm{d}t = \frac{\mathbb{E}_{p(\boldsymbol{x}_0|\boldsymbol{x}_t)}\left[\exp\left(\frac{-\Phi(\boldsymbol{x}_0)}{\alpha}\right)\mathrm{d}\bar{\mathbf{w}}\right]}{\mathbb{E}_{p(\boldsymbol{x}_0|\boldsymbol{x}_t)}\left[\exp\left(\frac{-\Phi(\boldsymbol{x}_0)}{\alpha}\right)\right]} \quad (8)$$

在此方程中，样本形式 $p(\boldsymbol{x}_0|\boldsymbol{x}_t)$ 由无控制过程获得，即方程 3。可以看出，方程 8 仍面临与方程 5 相同的挑战：计算期望项需要模拟完整轨迹以获得 $\boldsymbol{x}_0$。SCG（Huang 等，2024年）采用一种直观的近似方法来估计最优策略。他们首先考虑如下极限：

$\alpha \to 0$，此时，公式 8 退化为选择使 $\Phi(\boldsymbol{x}_0)$ 最小的 $\mathrm{d}\bar{\mathbf{w}}$。接下来，他们采用与 DPS 类似的方法，利用单步去噪结果 $\Phi(\hat{\boldsymbol{x}}_{0|t})$ 来估计 $\Phi(\boldsymbol{x}_0)$。实际操作中，SCG 会从 DDIM 求解器（Song、Meng 和 Ermon 2020年）的扩散反向核 $\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^n \sim p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}) = \mathcal{N}(\mu_t, \sigma_t^2 I)$ 中采样多个候选项，并通过启发式方法保留其干净估计最接近 $\boldsymbol{y}$ 的粒子作为下一状态。其他粒子则被丢弃。

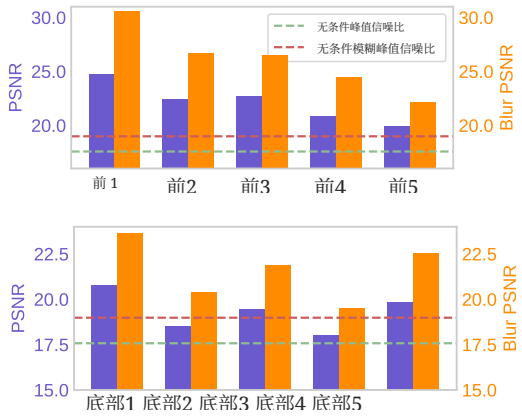


图2：SCG 中针对不同粒子选择方式的黑洞成像实证结果。上方展示了选择排名靠前的若干粒子所获得的结果；下方展示了选择排名靠后的若干粒子并反转其噪声符号所产生的效果。

方法

本节首先通过实证分析揭示 SCG 方法的低效之处。在此基础上，我们提出受约束粒子搜索（***CPS***）框架，该框架利用所有可用的候选位置信息来识别最优粒子。最后，我们说明如何将 CPS 与重启策略结合，以减少累积误差并进一步提升解的稳健性。

利用所有粒子的信息

尽管概念上简单，SCG（Huang 等人，2024）等方法在反向去噪过程中会丢弃除一个候选粒子之外的所有粒子，因此粒子效率不佳。我们推测，这一过程中会损失大量有价值的信息，因为大多数粒子都包含有助于引导去噪轨迹的见解。为检验这一假设，我们使用黑洞成像逆问题（Zheng 等人，2025b）开展了一项实证研究。

我们的结果如图 2（上）所示，为这一假设提供了有力证据。排名第一的粒子不仅能有效引导去噪过程朝观测信号的方向推进，排名第二和第三的粒子也发挥了显著作用，且表现均优于无条件采样。此外，我们还发现了一个令人惊讶的结果：即使是表现最差的五个粒子（排名倒数第一至第五），只要反转其噪声符号，仍能引导去噪过程朝正确方向推进，如图 2（下）所示。

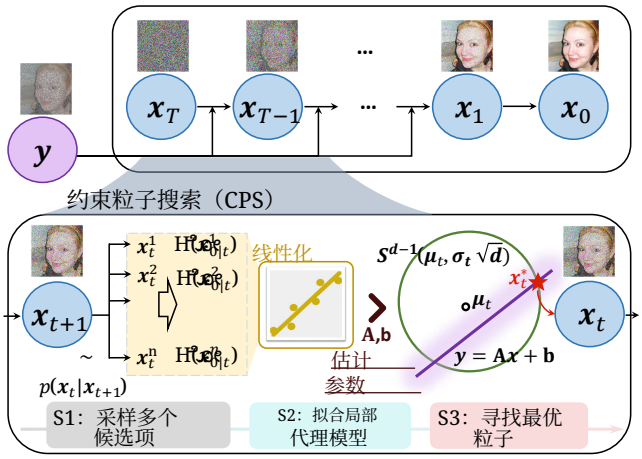


图3：用于逆问题求解的约束粒子搜索（CPS）概览。在每个时间步，该过程包含三个步骤。步骤 *1*：采样多个候选粒子；步骤 *2*：使用这些样本拟合前向过程替代模型；步骤 *3*：利用替代模型确定超球面上的最优粒子，作为下一个状态。

这些发现有力地表明，引导扩散过程所需的信息并非仅存在于单个粒子中，而是分布在整个集合中。受这些实证观测的启发，我们提出了一种新方法，以系统地利用所有候选项中的信息。

约束粒子搜索（CPS）

SCG方法可以理解为先从无条件转移核 $p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1})$中采样，然后被动地选择最有希望的候选项。相比之下，我们采用更主动的方法，在寻找有希望的粒子的同时，加入与无条件先验高密度区域相契合的额外约束。我们将这一思路重新表述为以下约束优化问题：

$$\boldsymbol{x}_t^* = \arg\min_{\boldsymbol{x}_t} \Phi\left(\hat{\boldsymbol{x}}_{0|t}\right) \quad \text{s.t. } \boldsymbol{x}_t \sim p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}) \tag{9}$$

这一优化框架更加灵活，使粒子搜寻更为有效，因为它不再局限于仅从采样集合中选择粒子。

求解这一优化问题面临两项主要挑战。首先，我们必须解决如何最小化目标函数，尤其是在处理黑箱观测过程时。其次，我们需要在实践中实现无条件核的约束。接下来，我们将概述应对这些挑战的方案。

观测过程的代理模型。优化终端代价通常需要估计其相对于$\boldsymbol{x}_t$的梯度，可按如下方式推导：

$$\nabla_{\boldsymbol{x}_t}\Phi\left(\hat{\boldsymbol{x}}_{0|t}\right) = \left(\frac{\partial \mathcal{H}(\hat{\boldsymbol{x}}_{0|t})}{\partial \boldsymbol{x}_t}\right)^\top \left(\mathcal{H}(\hat{\boldsymbol{x}}_{0|t}) - \boldsymbol{y}\right) \tag{10}$$

然而，计算项 $\partial\mathcal{H}(\hat{\boldsymbol{x}}_{0|t})/\partial\boldsymbol{x}_t$ 颇具挑战，因为我们无法直接获取梯度 $\mathcal{H}(\cdot)$。为了解决这一问题，我们采用统计线性化（Roberts 和 Spanos 2003），在由 $p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1})$ 定义的区域附近估计一个局部替代模型。具体而言，我们首先像 SCG 中那样采样 n 个粒子 $\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^n \sim p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1})$，并估计相应的值 $\mathcal{H}\left(\hat{\boldsymbol{x}}_{0|t}^1\right), \cdots, \mathcal{H}\left(\hat{\boldsymbol{x}}_{0|t}^n\right)$。然后，我们用参数 $(\mathbf{A},\mathbf{b})$ 将这些值拟合为最佳线性模型。这将得到以下优化问题：

$$\min_{\mathbf{A},\mathbf{b}} \mathbb{E}_{\boldsymbol{x}_t\sim p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1})} \|\mathcal{H}\left(\hat{\boldsymbol{x}}_{0|t}\right) - (\mathbf{A}\boldsymbol{x}_t + \mathbf{b})\|^2 \tag{11}$$

该问题的闭式解为：

$$\begin{aligned} \mathbf{A} &= \mathrm{Cov}(\mathcal{H}\left(\hat{\boldsymbol{x}}_{0|t}\right), \boldsymbol{x}_t)\mathrm{Cov}(\boldsymbol{x}_t, \boldsymbol{x}_t)^{-1} \\ \mathbf{b} &= \mathbb{E}[\mathcal{H}\left(\hat{\boldsymbol{x}}_{0|t}\right)] - \mathbf{A}\mathbb{E}[\boldsymbol{x}_t] \end{aligned} \tag{12}$$

鉴于扩散模型的转移核是各向同性高斯密度，我们分别将其均值和方差表示为 μ_t 和 σ_t^2。这样便可进一步简化表达式。接下来，我们使用采样粒子 $\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^n$进行经验估计，得到 $\mathbf{A}$ 和 $\mathbf{b}$的如下表达式，

$$\mathbf{A} = \frac{1}{n\sigma_t^2}\sum_{i=1}^{n}(\mathcal{H}\left(\hat{\boldsymbol{x}}_{0|t}^i\right) - \bar{\mathcal{H}})(\boldsymbol{x}_t^i - \mu_t)^\top \tag{13}$$

$$\mathbf{b} = \bar{\mathcal{H}} - \mathbf{A}\mu_t$$

其中，$\bar{\mathcal{H}}$ 表示 $\frac{1}{n}\sum_{i=1}^{n}\mathcal{H}\left(\hat{\boldsymbol{x}}_{0|t}^i\right)$，为简洁起见。在这一阶段，我们可以使用代理模型替代终端代价中不可微的部分。此外，我们注意到，式 13 中的$\mathbf{A}$ 实际上是采样粒子的加权线性组合，整合了所有候选信息。这使我们能够选出优于任何单个样本的粒子，从而更有效地识别下一状态。

转移核约束。在上一节中，我们介绍了使用线性代理来估计终端代价，从而优化目标函数。式 9 还要求优化变量从无条件分布中采样，以确保代理模型处于可信范围内，并确保模型在推理过程中不会偏离基础模型的先验。

转移核 $p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1})$ 是一个高斯分布，其对数似然函数易于计算，并且倾向于将 $\boldsymbol{x}_t$ 推向均值。正如 Menon 等人（2020年）所示，来自 d 维高斯分布的样本，其范数服从一种 χ^d 分布。对于扩散模型中常见的高维空间，该 χ^d 分布的概率质量高度集中在数值 $\sigma_t\sqrt{d}$ 附近，形成一个超球面 $S^{d-1}(\mu_t, \sigma_t\sqrt{d})$（Samuel 等人，2023年）。扩散模型通常偏好符合基于范数的分布特征的输入数据，而不是集中在均值附近。因此，我们旨在使 $\boldsymbol{x}_t$ 保持在 $p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1})$ 的高概率质量区域内。基于此，原始优化问题（式 9）可以借助代理模型解释如下：

$$\min_{\boldsymbol{x}_t} \|\boldsymbol{y} - (\mathbf{A}\boldsymbol{x}_t + \mathbf{b})\|^2 \quad \text{s.t. } \boldsymbol{x}_t \in S^{d-1}(\mu_t, \sigma_t\sqrt{d}) \tag{14}$$

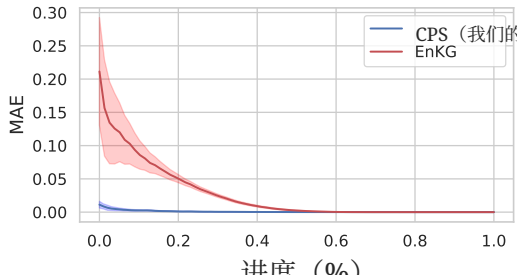


图4：4×超分辨率任务中，CPS 与 EnKG 在 FFHQ 数据集上的詹森间隙演变。

上述约束优化问题可以使用拉格朗日乘数法求解（Bertsekas 2014），从而得到最优解对应的以下方程组：

$$\begin{cases} \mathbf{A}^\top(\mathbf{A}\boldsymbol{x}_t + \mathbf{b} - \boldsymbol{y}) + \lambda(\boldsymbol{x}_t - \mu_t) = 0 \\ \|\boldsymbol{x}_t - \mu_t\|^2 - \sigma_t^2 d = 0 \end{cases} \tag{15}$$

其中，λ 是拉格朗日乘数。直接求解该方程组需要进行矩阵求逆和求根，这在高维空间中颇具挑战。我们注意到，反向核的 σ_t 通常较小，因此在寻找最优解时，约束条件会占据主导地位。因此，乘数 λ 预计会足够大。在这些条件下，我们可以解析地推导出最优解的渐近形式：

$$\boldsymbol{x}_t^* \approx \mu_t + \sigma_t\sqrt{d}\frac{\mathbf{A}^\top(\boldsymbol{y} - \bar{\mathcal{H}})}{\|\mathbf{A}^\top(\boldsymbol{y} - \bar{\mathcal{H}})\|} \tag{16}$$

详细推导见附录 A。最终，解 $\boldsymbol{x}_t^*$ 可视为结合所有候选信息合成的最优粒子。相比之下，SCG 仅选择终端代价最小的粒子，而我们的方法更具理论依据，也更高效。图3概述了我们的方法。

请注意，EnKG（Zheng 等人，2025a）采用了类似的线性化策略。然而，EnKG 在整个全局空间上执行线性化，这可能导致不准确的近似，并在求逆协方差矩阵（公式 12）时产生困难。相比之下，我们的方法围绕 $p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1})$进行局部线性化。在图 4 中，我们使用采样粒子计算函数 $\mathcal{H}(\mathbb{E}[\boldsymbol{x}_0|\cdot])$ 的詹森间隙（Gao、Sitharam 和 Roitberg 2017）的平均绝对误差（MAE），以评估其偏离线性行为的程度。结果表明，EnKG 在生成的早期阶段会产生显著误差，而 CPS 始终保持较小的误差。此外，CPS 通过合成候选项，在每一步寻找一个最优粒子，而不像 EnKG 那样显式维护完整的粒子集合。我们在实验部分进行的消融实验表明，CPS 的效率显著高于后者生成的粒子。

通过重启提升鲁棒性

实际上，我们观察到，不同的初始噪声条件会导致图像复原任务产生不同的结果。

算法 1：CPS 的采样过程

需要： 观测 $\boldsymbol{y}$，初始噪声 $\boldsymbol{x}_T \sim \mathcal{N}(0, I)$，预训练的扩散模型 $p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}) = \mathcal{N}(\mu_t, \sigma_t I)$，数据维度 d，以及重启次数 N_r

对于 $t = T-1, \cdots, 0$ **do**
  对于 $r = N_r, \cdots, 1$ **do**
    ▷ 从反向核中采样
    $\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^n \sim p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1})$
    ▷ 拟合前向过程替代模型
    $\mathbf{A} \leftarrow \frac{1}{n\sigma_t^2}\sum_{i=1}^n (\mathcal{H}(\hat{\boldsymbol{x}}_{0|t}^i) - \bar{\mathcal{H}})(\boldsymbol{x}_t^i - \mu_t)^\top$
    ▷ 将最优粒子求解为下一状态
    $\boldsymbol{x}_t \leftarrow \mu_t + \sigma_t\sqrt{d}\frac{\mathbf{A}^\top(\boldsymbol{y} - \bar{\mathcal{H}})}{\|\mathbf{A}^\top(\boldsymbol{y} - \bar{\mathcal{H}})\|}$
    如果 $r > 1$ **那么**
      ▷ 为重新启动重新添加噪声
      $\boldsymbol{x}_{t+1} \leftarrow p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)$
    结束条件判断
  结束循环
结束循环
确保： $\boldsymbol{x}_0$

当初始噪声不理想时，需要进行大幅调整才能纠正整体不匹配。此外，Tweedie 公式的一步估计可能会引入近似误差，从而影响性能。

为应对这些挑战，我们引入了近期提出的重启策略（Xu 等人，2023；Lugmayr 等人，2023）。该方法提供了一种有效途径，可减轻采样过程中的误差累积。该过程是一个简单的两步迭代：在时间步 t，根据公式 2 对样本重新加噪，然后将其返回至前一个时间步 $t+1$。随后，这个重新加噪的样本会与 CPS 一起用于引导去噪。这种逐步处理的方法能够逐渐修正累积误差，从而增强 CPS 的鲁棒性，并确保获得更可靠的解。包含重启策略在内的完整采样过程详见算法 1。

实验

为验证我们方法的有效性，我们在图像逆问题和两个科学逆问题上进行了实验。我们的基线包括若干最近提出的用于求解逆问题的无梯度方法，例如 SCG（Huang 等人，2024）、DPG（Tang 等人，2024年）和 EnKG（Zheng 等人，2025a）。所有无梯度方法均使用相当数量的前向传播进行评估，以确保公平比较。此外，我们还纳入了先进的基于梯度的方法，用于线性图像问题的比较，包括 DDRM（Kawar 等人 2022a）、DPS（Chung 等人 2022）、ΠGDM（Song 等人 2023）、RED-diff（Mardani 等，2023）和 DAPS（Zhang 等人 2025）。所测试逆问题的公式详见附录 B.1。更多实现细节见附录 B.2。

	方法	图像修复（95%）			超分辨率（×4）			去模糊（高斯）			JPEG（质量因子=5）		
		峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓
梯度-based	DDRM	21.65	0.693	0.270	28.02	0.828	0.213	24.69	0.753	0.187	-	-	-
	DPS	22.49	0.701	0.236	27.65	0.822	0.120	25.59	0.790	0.136	-	-	-
	ΠGDM	23.13	0.774	0.231	27.25	0.829	0.135	26.76	0.812	0.141	18.25	0.593	0.304
	RED-diff	24.43	**0.807**	0.222	26.85	0.835	0.191	28.01	**0.853**	0.164	-	-	-
	DAPS	22.71	0.754	0.226	**28.53**	**0.840**	0.132	**28.70**	0.843	0.133	-	-	-
梯度-free	SCG	6.15	0.321	0.807	6.16	0.319	0.805	7.01	0.322	0.792	6.17	0.362	0.825
	DPG	19.83	0.651	0.299	24.81	0.758	0.187	24.11	0.740	0.196	22.52	0.734	0.246
	EnKG	22.64	0.767	0.286	21.72	0.737	0.293	20.84	0.731	0.334	20.89	0.704	0.257
	CPS（我们的方法）	**24.90**	0.794	**0.161**	27.70	0.834	**0.115**	27.21	0.822	**0.128**	**23.23**	**0.784**	**0.223**

表1：不同方法在 FFHQ 上的定量结果，包括三个线性逆问题和一个不可微逆问题。最佳结果以**第一名**、第二名和第三名标示。对于不适用的[illegible]-”标[illegible]

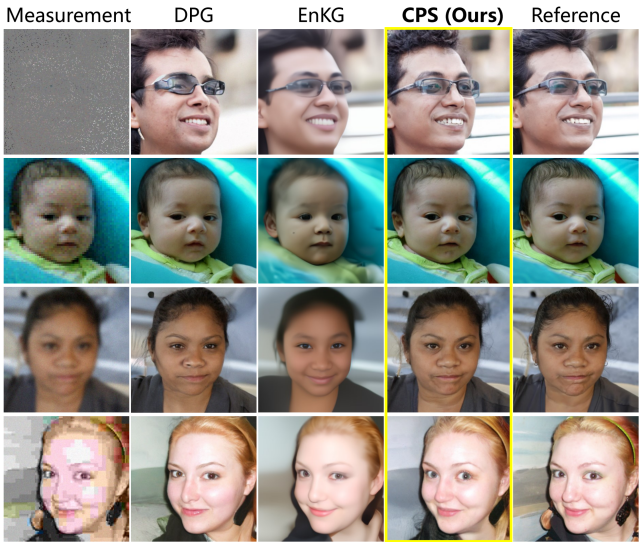


图5：FFHQ 上图像逆问题的定性结果：从上到下依次为图像修复、超分辨率、去模糊和 JPEG复原。

图像逆问题

实验设置：我们首先在图像逆问题上评估我们的方法。我们从 FFHQ 数据集（Karras、Laine 和 Aila，2019年）中采样 100 张图像用于验证，并使用 Chung 等人（2022年）提供的检查点。所有图像均被调整为 256 × 256像素，并使用三项标准图像复原指标评估其质量：峰值信噪比、结构相似性指数和学习感知图像块相似度（Zhang 等人，2018）。

我们聚焦于三个线性逆问题：95% 缺失的随机掩码补全、 4×超分辨率，以及使用标准差为 3（61× 61）的高斯模糊核进行去模糊。为评估无梯度方法的实用性，我们还纳入了 JPEG复原任务（Song 等人 2023），其中前向模型不可微。对于 JPEG复原，我们采用质量因子（质量因子）为 5 的设置。所有测量均加入了$\sigma_y = 0.05$的加性高斯噪声。

结果：表1和图5分别展示了定量结果和可视化结果。我们的发现表明，SCG

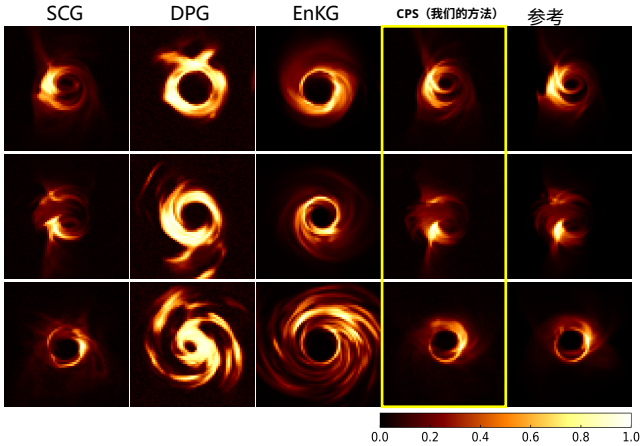


图6：黑洞成像的定性结果。

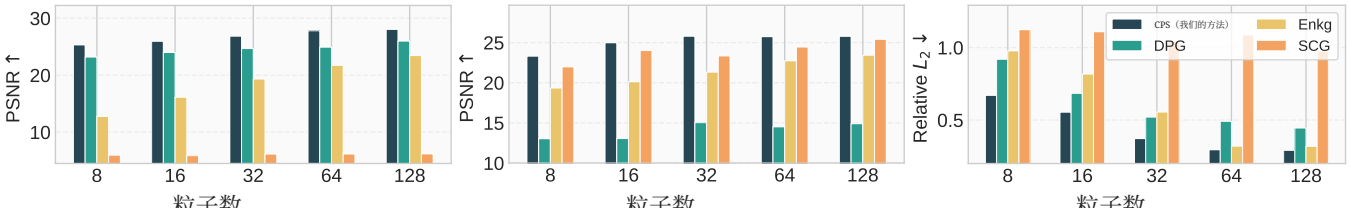
无法有效处理图像逆问题。相比之下，CPS 的表现显著优于无梯度方法，并且与基于梯度的方法不相上下。据我们所知， ΠGDM 是唯一一种能够通过构建伪逆来进行 JPEG 复原的方法（Song 等人 2023）。然而，我们观察到， ΠGDM 在有噪场景中表现不佳，性能明显较差。这凸显了进一步开发无梯度反演方法的重要性。

方法	100%		10%		3%	
	峰值信噪比 ↑	BPSNR ↑	峰值信噪比 ↑	BPSNR ↑	峰值信噪比 ↑	BPSNR ↑
SCG	25.45	30.56	24.97	31.45	24.65	31.28
DPG	14.51	16.75	13.57	16.35	12.16	13.71
EnKG	23.03	25.67	22.77	25.18	21.41	25.14
CPS（本文）	**25.74**	**31.84**	**25.34**	**31.54**	**25.03**	**31.81**

表2：无梯度方法在黑洞成像问题上的比较，包括三种观测时间比例：100%、10%和3%。

黑洞成像

实验设置：由于观测稀疏，黑洞成像呈现出高度非线性且不适定的前向模型。测量通过甚长基线



(a) 4× 图像超分辨率　(b) 黑洞成像　（c）流体数据同化

图7：三项任务中无梯度方法的粒子效率对比。

干涉测量（VLBI）获得，其中每个望远镜对在特定时间提供可见度数据。为减少大气湍流和热噪声带来的误差，通常将多个可见度数据点组合为闭合相位 和对数闭合振幅 （Blackburn 等人 2020年）。除这些约束外，黑洞的总流量 也常被纳入似然函数。在评估中，我们假设对前向模型具有黑盒访问。我们采用 Zheng 等人（2025b）的预训练扩散模型，并在测试集的 100 张调整尺寸后的 64×64 图像上对其进行评估。我们考虑三种观测时间比例设置：100%、10% 和 3%。重建质量通过计算重建图像与真实值之间的峰值信噪比来衡量，同时计算模糊峰值信噪比（BPSNR）；后者使用滤波器进行平滑，以评估低频重建性能。

结果： 图6比较了使用无梯度方法重建的黑洞图像。CPS 的结果在视觉上与真实值更加一致。表2展示了定量比较结果，其中 CPS 的表现最佳，SCG 紧随其后。相比之下，面对这类高度非线性问题时，DPG 的表现较差。

	×2	×4	×8
方法	相对 $L_2\downarrow$	相对 $L_2\downarrow$	相对 $L_2\downarrow$
SCG	1.087	1.095	1.123
DPG	0.491	0.592	0.836
EnKG	0.320	0.528	0.821
CPS（我们的方法）	**0.295**	**0.469**	**0.684**

表3：流体数据同化中无梯度方法的比较，包括三种类型的稀疏观测：2×、 4×和 8×下采样。

流体数据同化

实验设置： 流体动力学通常由 Navier–Stokes（NS）方程组建模，这是一组非线性偏微分方程（Chandler 和 Kerswell，2013）。在本实验中，我们针对 $T = 1$ 处稀疏且含噪的涡量测量进行流体同化，目标是恢复二维初始涡量场。鉴于偏微分方程数值求解器可能需要数千个离散时间步，通过自动微分获取准确的数值导数面临重大挑战。因此，这一问题非常适合用于验证无梯度方法。我们使用 Zheng 等人提供的预训练检查点。

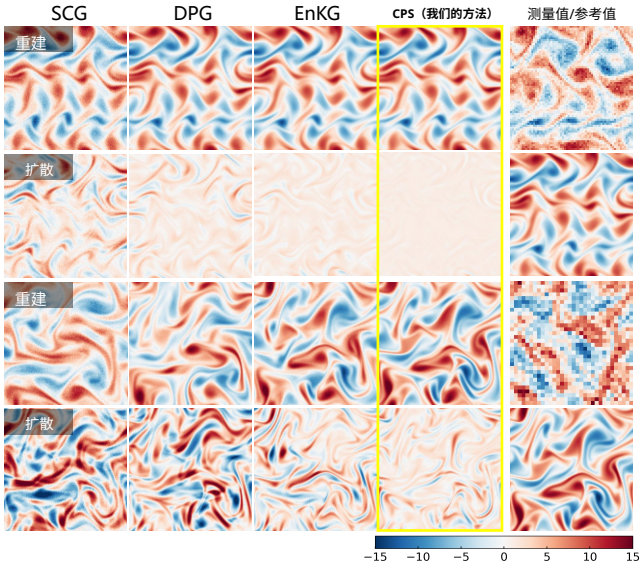


图8：恢复初始涡量场的定性结果。奇数行显示重建结果，偶数行显示与参考结果的差异。

（2025b）以及 10 个尺寸为 128×128 的涡量场进行测试。结果采用相对 L_2 误差进行评估。我们考虑三种设置来模拟稀疏测量，包括 2×、 4× 和 8× 下采样。所有测量均含有噪声，噪声水平为 $\sigma_y = 2.0$。

结果： 图8展示了一个简单示例和一个具有挑战性的可视化结果。在简单示例中，所有方法都能够有效恢复整体结构。然而，在更困难的场景中，CPS通过从稀疏含噪测量中恢复初始涡量场模式，显著优于基线方法。SCG和DPG在这些条件下常常失效。EnKG的表现次之，但在一些局部区域存在明显误差。表3中的定量结果进一步支持了我们的发现。

消融研究

粒子效率比较： 无梯度方法通常需要对多个粒子进行采样。在本研究中，我们使用 8、16、32、64 和 128 个粒子来评估这些方法的粒子效率。如图 7 所示，与基线方法相比，CPS 的粒子效率显著更高，并且在各种任务中都保持了稳

定性和竞争力。即使只有 8 个粒子，CPS 也表现出色。相比之下，EnKG 难以应对较小的粒子集，需要更多粒子才能有效运行。DPG 在高度病态的黑洞成像问题上表现困难，而 SCG 则无法解决图像相关问题。

重启的有效性：我们通过使用不同的随机种子初始化噪声，研究了重启策略对 CPS 的影响。如附录 C 所示，重启策略不仅提升了整体性能，还增强了算法的稳定性和鲁棒性。它有效解决了从较差初始噪声中采样的问题，避免了反演过程中的失败。

结论

在本研究中，我们提出了***CPS***，一种用于借助扩散模型解决逆问题的新型无梯度方法。CPS 利用所有候选粒子的信息，并根据先验的高密度区域施加约束，从而实现高效的粒子搜索。实验结果表明，CPS 优于其他无梯度方法，且取得了与基于梯度的方法相近的结果。这些结果展示了 CPS 在无法获取正向过程的梯度时，作为可靠解决方案的潜力。

致谢

本研究得到国家自然科学基金资助，项目编号为 62405014、62325101 和 62031001。

参考

Bertsekas, D. P. 2014. 约束优化与拉格朗日乘子方法. Academic Press. Blackburn, L.; Pesce, D. W.; Johnson, M. D.; Wielgus, M.; Chael, A. A.; Christian, P.; and Doeleman, S. S. 2020年. 干涉测量数据中的闭合统计量. 《天体物理学杂志》, 894(1): 31. Broyden, C. G. 1965. 求解非线性联立方程的一类方法. 《计算数学》, 19(92): 577–593. Chandler, G. J.; and Kerswell, R. R. 2013. 嵌入湍流二维 Kolmogorov 流中的不变循环解. 《流体力学杂志》, 722: 554–595. Chung, H.; Kim, J.; Mccann, M. T.; Klasky, M. L.; and Ye, J. C. 2022年. 面向一般含噪逆问题的扩散后验采样. *arXiv* 预印本 *arXiv:2209.14687*. Daniel, J. W. 1967. 线性与非线性算子方程的共轭梯度法. 《*SIAM* 数值分析杂志》, 4(1): 10–26. Duong, C. N.; Truong, T.-D.; Luu, K.; Quach, K. G.; Bui, H.; and Roy, K. 2020年. Vec2face：从人脸识别中的黑盒特征还原人脸. 载于 *IEEE/CVF* 计算机视觉与模式识别会议论文集, 6132–6141. Efron, B. 2011年. Tweedie 公式与选择偏差. 《美国统计学会杂志》, 106(496): 1602–1614. Ehrlich, M.；Davis, L.；Lim, S.-N.；以及 S hrivastava, A. 2020年。量化引导的 JPEG 伪影校正。载于欧洲计算机视觉会议，293–309。Springer。Gao, X.；Sitharam, M.；以及 Roitberg, A. E. 2017年。关于詹森间隙的界及其对均值集中分布的含义。*arXiv* 预印本 *arXiv:1712.05267*。Gillhofer, M.；Ramsauer, H.；Brandstetter, J.；Sch¨afl, B.；以及 Hochreiter, S. 2019。基于生成对抗网络的黑盒逆问题求解器。载于 *NeurIPS 2019* 使用深度网络求解逆问题研讨会。Ho, J.；Jain, A.；以及 Abbeel, P. 2020年。去噪扩散概率模型。神经信息处理系统进展，33：6840–6851。Huang, Y.；Ghatare, A.；Liu, Y.；Hu, Z.；Zhang, Q.；Sastry, C. S.；Gururani, S.；Oore, S.；以及 Yue, Y. 2024年。结合不可微规则引导扩散的符号音乐生成。载于国际机器学习会议，19772–19797。PMLR。Iordache, M.-D.；Bioucas-Dias, J. M.；以及 Plaza, A. 2012年。稀疏高光谱解混的全变分空间正则化。*IEEE* 地球科学与遥感汇刊，50(11)：4484–4502。Karras, T.；Laine, S.；以及 Aila, T. 2019。基于风格的生成对抗网络生成器架构。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，4401–4410。Katzfuss, M.；Stroud, J. R.；以及 Wikle, C. K. 2016年。理解集合卡尔曼滤波器。美国统计学家，70(4)：350–357。Kawar, B.；Elad, M.；Ermon, S.；以及 Song, J. 2022a。去噪扩散复原模型。神经信息处理系统进展，35：23593–23606。Kawar, B.；Song, J.；Ermon, S.；以及 Elad, M. 2022年b。使用去噪扩散复原模型进行 JPEG 伪影校正。*arXiv* 预印本 *arXiv:2209.11888*。Li, Z.；Kovachki, N.；Azizzadenesheli, K.；Liu, B.；Bhattacharya, K.；Stuart, A.；以及 Anandkumar, A. 2020年。用于参数化偏微分方程的傅里叶神经算子。*arXiv* 预印本 *arXiv:2010.08895*。Lipman, Y.；Chen, R. T.；Ben-Hamu, H.；Nickel, M.；以及 Le, M. 2022年。用于生成建模的流匹配。*arXiv* 预印本 *arXiv:2210.02747*。Liu, G.；Sun, H.；Li, J.；Yin, F.；以及 Yang, Y. 2023年。通过捷径采样加速用于逆问题的扩散模型。*arXiv* 预印本 *arXiv:2305.16965*。Lugmayr, A.；Danelljan, M.；Romero, A.；Yu, F.；Timofte, R.；以及 Van Gool, L. R. 2023年。使用去噪扩散概率模型进行图像修复。载于 *IEEE/CVF* 计算机视觉与模式识别会议论文集，11461–11471。Mardani, M.；Song, J.；Kautz, J.；以及 Vahdat, A. 2023年。从变分视角使用扩散模型求解逆问题。*arXiv* 预印本 *arXiv:2305.04391*。McCann, M. T.；Jin, K. H.；以及 Unser, M. 2017。用于成像逆问题的卷积神经网络：综述。*IEEE* 信号处理杂志，34(6)：85–95。

Menon, S.；Damian, A.；Hu, S.；Ravi, N.；以及 Rudin, C. 2020年。Pulse：通过生成模型的潜在空间探索进行自监督图像上采样。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，2437–2445。Oliver, D. S.；Reynolds, A. C.；以及 Liu, N. 2008。石油储层表征与历史拟合的反演理论。Pavon, M. 1989年。随机控制与非平衡热力学系统。应用数学与优化，19(1)：187–202。Roberts, J. B.；以及 Spanos, P. D. 2003。随机振动与统计线性化。Courier Corporation。Rombach, R.；Blattmann, A.；Lorenz, D.；Esser, P.；以及 Ommer, B. 2022年。使用潜在扩散模型进行高分辨率图像合成。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，10684–10695。Saad, Y.；以及 Schultz, M. H. 1986。GMRES：求解非对称线性系统的广义最小残量算法。*SIAM* 科学与统计计算杂志，7(3)：856–869。Samuel, D.；Ben-Ari, R.；Darshan, N.；Maron, H.；以及 Chechik, G. 2023年。用于文生图的范数引导潜在空间探索。神经信息处理系统进展，36：57863–57875。Schwarzbach, C.；B¨orner, R.-U.；以及 Spitzer, K. 2011年。用于地球电磁学的三维自适应高阶有限元模拟——海洋可控源电磁法示例。国际地球物理学杂志，187(1)：63–74。Song, J.；Meng, C.；以及 Ermon, S. 2020年。去噪扩散隐式模型。*arXiv* 预印本 *arXiv:2010.02502*。Song, J.；Vahdat, A.；Mardani, M.；以及 Kautz, J. 2023年。用于逆问题的伪逆引导扩散模型。载于国际学习表征会议。Song, Y.；Shen, L.；Xing, L.；以及 Ermon, S. 2021年。使用基于得分的生成模型求解医学成像中的逆问题。*arXiv* 预印本 *arXiv:2111.08005*。Song, Y.；Sohl-Dickstein, J.；Kingma, D. P.；Kumar, A.；Ermon, S.；以及 Poole, B. 2020年。通过随机微分方程实现基于得分的生成建模。*arXiv* 预印本 *arXiv:2011.13456*。Sun, Y.；Wu, Z.；Chen, Y.；Feng, B. T.；以及 Bouman, K. L. 2024年。使用基于得分的生成先验实现可证明的概率成像。*IEEE* 计算成像汇刊。Tang, H.；Xie, T.；Feng, A.；Wang, H.；Zhang, C.；以及 Bai, Y. 2024年。通过后验采样求解一般噪声逆问题：策略梯度视角。载于国际人工智能与统计会议，2116–2124。PMLR。Theodorou, E.；Buchli, J.；以及 Schaal, S. 2010年。强化学习的广义路径积分控制方法。机器学习研究期刊，11：3137–3181。Tibshirani, R. 1996年。通过套索进行回归收缩与选择。英国皇家统计学会杂志*B*辑：统计方法学，58(1): 267–288。Tonolini, F.; Radford, J.; Turpin, A.; Faccio, D.; and Murray- Smith, R. 2020年。计算成像逆问题的变分推断。机器学习研究期刊，21(179): 1–46。Vincent, P. 2011年。评分匹配与去噪自编码器之间的联系。神经计算，23(7): 1661– 1674。Wang, Y.; Yu, J.; and Zhang, J. 2022年。使用去噪扩散零空间模型进行零样本图像复原。*arXiv* 预印本 *arXiv:2212.00490*。Williams, G.; Aldrich, A.; and Theodorou, E. A. 2017年。模型预测路径积分控制：从理论到并行计算。引导、控制与动力学杂志，40(2): 344–357。Xu, Y.; Deng, M.; Cheng, X.; Tian, Y.; Liu, Z.; and J aakkola, T. 2023年。通过重启采样改进生成过程。神经信息处理系统进展，36: 76806–76838。Xue, S.; Yi, M.; Luo, W.; Zhang, S.; Sun, J.; Li, Z.; and Ma, Z.-M. 2023年。SA-Solver：用于扩散模型快速采样的随机亚当斯求解器。神经信息处理系统进展，36: 77632–77674。Zhang, B.; Chu, W.; Berner, J.; Meng, C.; Anandkumar, A.; and Song, Y. 2025年。通过解耦噪声退火改进扩散逆问题求解。载于计算机视觉与模式识别会议论文集，20895–20905。Zhang, Q.; and Chen, Y. 2023年。路径积分采样器：一种用于采样的随机控制方法。载于国际学习表征会议。Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018年。深度特征作为感知度量的惊人效力。载于*IEEE*计算机视觉与模式识别会议论文集，586–595。Zheng, H.; Chu, W.; Wang, A.; Kovachki, N. B.; Baptista, R.; and Yue, Y. 2025a。一种用于逆问题的无导数方法：集成卡尔曼扩散引导。机器学习研究汇刊。Zheng, H.; Chu, W.; Zhang, B.; Wu, Z.; Wang, A.; Feng, B.; Zou, C.; Sun, Y.; Kovachki, N. B.; Ross, Z. E.; et al. 2025b。InverseBench：物理科学逆问题中即插即用扩散先验的基准测试。载于第十三届国际学习表征会议。

附录 A：推导 A1：正文公式 16 的推导

拉格朗日乘数法：对于正文公式 14 中给出的球面优化问题，我们旨在求解：

$$\min_{\boldsymbol{x}_t} \|\boldsymbol{y}-(\mathbf{A}\boldsymbol{x}_t+\mathbf{b})\|^2 \quad \text{s.t. } \|\boldsymbol{x}_t-\mu_t\|^2=\sigma_t^2 d \tag{17}$$

我们可以使用拉格朗日乘数法放宽该问题来求解。拉格朗日函数为：

$$L=\|\boldsymbol{y}-(\mathbf{A}\boldsymbol{x}_t+\mathbf{b})\|^2+\lambda(\|\boldsymbol{x}_t-\mu_t\|^2-\sigma_t^2 d) \tag{18}$$

令关于变量$\boldsymbol{x}_t$和乘数 λ 的偏导数为零，可得到以下用于求取最优解的方程组：

$$\begin{cases} \mathbf{A}^\top(\mathbf{A}\boldsymbol{x}_t+\mathbf{b}-\boldsymbol{y})+\lambda(\boldsymbol{x}_t-\mu_t)=0 \\ \|\boldsymbol{x}_t-\mu_t\|^2-\sigma_t^2 d=0 \end{cases} \tag{19}$$

求解该方程组即可找到最优候选粒子。然而，通常不存在解析解，因此我们一般会采用数值方法。具体而言，第一个方程可以改写为关于 $\boldsymbol{x}_t$ 的线性方程组：

$$(\mathbf{A}^\top\mathbf{A}+\lambda I)\boldsymbol{x}_t=\mathbf{A}^\top(\boldsymbol{y}-\mathbf{b})+\lambda\mu_t \tag{20}$$

该方程组通常使用迭代方法求解，例如共轭梯度（CG）法（Daniel 1967）或 GMRES（Saad and Schultz 1986）。获得 $\boldsymbol{x}_t(\lambda)$ 的估计值后，将其代入第二个方程，随后可使用求根算法（例如牛顿法（Broyden 1965））求解 λ：

$$\begin{aligned} &\|\boldsymbol{x}_t(\lambda)-\mu_t\|^2-\sigma_t^2 d=0 \\ \text{s.t.}\quad &\boldsymbol{x}_t(\lambda)=(\mathbf{A}^\top\mathbf{A}+\lambda I)^{-1}[\mathbf{A}^\top(\boldsymbol{y}-\mathbf{b})+\lambda\mu_t] \end{aligned} \tag{21}$$

重复这两个步骤，我们便可以迭代收敛到一个解。然而，这种同时涉及求解线性方程组和求根的迭代方法，计算开销可能很大，而且数值上不稳定。因此，接下来我们提出一种近似闭式解，以规避这些问题。

近似解：我们首先对 $\boldsymbol{x}_t(\lambda)$ 的表达式进行一种“巧妙”的变换：

$$\begin{aligned} \boldsymbol{x}_t(\lambda)&=(\mathbf{A}^\top\mathbf{A}+\lambda\mathbf{I})^{-1}[\mathbf{A}^\top(\boldsymbol{y}-\mathbf{b})+\lambda\mu_t] \\ &=(\mathbf{A}^\top\mathbf{A}+\lambda\mathbf{I})^{-1}[\mathbf{A}^\top(\boldsymbol{y}-(\mathbf{A}\mu_t+\mathbf{b}))+\mathbf{A}^\top\mathbf{A}\mu_t+\lambda\mu_t] \\ &=(\mathbf{A}^\top\mathbf{A}+\lambda\mathbf{I})^{-1}[\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b})] \\ &\qquad+(\mathbf{A}^\top\mathbf{A}+\lambda\mathbf{I})^{-1}(\mathbf{A}^\top\mathbf{A}+\lambda\mathbf{I})\mu_t \\ &=(\mathbf{A}^\top\mathbf{A}+\lambda\mathbf{I})^{-1}\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b})+\mu_t \end{aligned} \tag{22}$$

接下来，我们利用这样一个事实：当 t 趋近于零时，DDIM 中去噪核的方差 σ_t 非常小，即 $\sigma_t\to 0$。要使式 21 成立，乘数 λ 必须足够大，即 $\lambda\to\infty$。在这一渐近区间内，项 $(\mathbf{A}^\top\mathbf{A}+\lambda\mathbf{I})^{-1}$ 可近似为 $\frac{1}{\lambda}\mathbf{I}$（因为 $\mathbf{A}^\top\mathbf{A}$ 与 $\lambda\mathbf{I}$ 相比可忽略不计）。因此，$\boldsymbol{x}_t(\lambda)$ 的渐近形式为：

$$\boldsymbol{x}_t(\lambda)\approx\frac{1}{\lambda}\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b})+\mu_t \tag{23}$$

将这一结果代回式 21 中的约束方程：

$$\left\|\left(\frac{1}{\lambda}\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b})+\mu_t\right)-\mu_t\right\|^2=\sigma_t^2 d \tag{24}$$

化简后得到：

$$\frac{1}{\lambda^2}\|\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b})\|^2=\sigma_t^2 d \tag{25}$$

求解乘子 λ：

$$\lambda=\frac{\|\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b})\|}{\sigma_t\sqrt{d}} \tag{26}$$

最后，将 λ 的这一表达式代回式 23，即可求得最优粒子的渐近形式：

$$\begin{aligned} \boldsymbol{x}_t^*&\approx\mu_t+\left(\frac{\sigma_t\sqrt{d}}{\|\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b})\|}\right)\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b}) \\ &=\mu_t+\sigma_t\sqrt{d}\frac{\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b})}{\|\mathbf{A}^\top(\boldsymbol{y}-\mathbf{A}\mu_t-\mathbf{b})\|} \\ &\quad(\text{Substituting } \mathbf{b}=\overline{\mathcal{H}}-\mathbf{A}\mu_t) \\ &=\mu_t+\sigma_t\sqrt{d}\frac{\mathbf{A}^\top(\boldsymbol{y}-\overline{\mathcal{H}})}{\|\mathbf{A}^\top(\boldsymbol{y}-\overline{\mathcal{H}})\|} \end{aligned} \tag{27}$$

结论得到了证实。

直观解释：这个近似解有着非常直观的几何解释：

1. 起点：该过程从约束球的中心 μ_t开始。

2. 方向：我们从起点 μ_t $\nabla_{\boldsymbol{x}_t}=$沿目标函数$\frac{1}{2}^T$的负梯度方向移动。目标函数的梯度为$\mathbf{A}(\mathbf{A}\boldsymbol{x}_t+\mathbf{b}-\boldsymbol{y})$。在$\boldsymbol{x}_t=\mu_t$处，负梯度方向为$\mathbf{A}^T(\boldsymbol{y}-(\mathbf{A}\mu_t+\mathbf{b}))$，即最速下降方向。

3. 步长：步长是一个固定距离 $\sigma_t\sqrt{d}$，由约束球的半径决定。

总之，当 σ_t 较小时，沿目标函数的最速下降方向移动固定距离 $\sigma_t\sqrt{d}$，即可近似得到最优解，起从球面中心出发 μ_t。

附录 B：实验设置

B1：逆问题的表述

图像修复。我们的图像修复问题的表述遵循近期研究（Chung 等人 2022；Song 等人 2023）。退化模型表示为：

$$\boldsymbol{y}=\mathbf{M}\odot\boldsymbol{x}+\boldsymbol{\eta} \tag{28}$$

其中$\mathbf{M}$是一个元素随机设为 0 或 1 的二值掩码，$\odot$表示逐元素乘法。目标是

图像修复任务是根据观测到的像素 y 恢复缺失的像素（即掩码为 0 的位置）。所有测量都包含加性高斯噪声 $\boldsymbol{\eta}$，其标准差为$\sigma_y = 0.05$。

超分辨率。 我们考虑如下退化模型下的双三次超分辨率（SR）：

$$\boldsymbol{y} = \boldsymbol{x}\downarrow_{sf}^{\text{bicubic}} + \boldsymbol{\eta} \tag{29}$$

其中，$\downarrow_{sf}^{\text{bicubic}}$表示采用指定缩放因子 sf进行双三次下采样。

图像去模糊。 含高斯噪声的图像去模糊线性退化模型表示为:

$$\boldsymbol{y} = \boldsymbol{x} \otimes \mathbf{k} + \boldsymbol{\eta} \tag{30}$$

其中，$\otimes$ 是一个二维卷积算子，$\mathbf{k}$ 是作用于所有图像通道的模糊核。

JPEG复原。 该问题表述如下:

$$\boldsymbol{y} = h(\boldsymbol{x}) + \boldsymbol{\eta} \tag{31}$$

其中，$h(\boldsymbol{x})$是 JPEG压缩函数。该函数包含离散余弦变换（DCT），随后进行量化。我们使用 Ehrlich等人（2020年）的量化矩阵来压缩原始 256×256 图像，并进行 2×2色度子采样。请注意，与 Song 等人（2023年）的设定不同，我们加入了额外的观测噪声，使这项任务更具挑战性，也更贴近现实。

黑洞成像。 黑洞成像测量通过甚长基线干涉测量（VLBI）技术获得（Blackburn 等人 2020年）。在这种技术中，每对望远镜（a，b）通过对源图像在时间 t 的特定空间傅里叶频率进行采样，获得复数可见度测量值 $V_{a,\,b}^{t}$。实际应用中，可见度数据会受到三类噪声的影响:望远镜增益误差、望远镜相位误差和基线热噪声。前两类噪声由大气湍流和仪器失准造成，可通过将多个可见度组合成能够消除噪声的数据产品来减轻：闭合相位和对数闭合振幅。其定义如下:

$$\begin{aligned}
\boldsymbol{y}^{\text{cph}}_{t,(a,b,c)} &= \angle\left(V^t_{a,b}V^t_{b,c}V^t_{a,c}\right) := \mathcal{H}^{\text{cph}}_{t,(a,b,c)}(\boldsymbol{x}) \\
\boldsymbol{y}^{\text{camp}}_{t,(a,b,c,d)} &= \log\left(\frac{|V^t_{a,b}||V^t_{c,d}|}{|V_{a,c}||V^t_{b,d}|}\right) := \mathcal{H}^{\text{camp}}_{t,(a,b,c,d)}(\boldsymbol{x})
\end{aligned} \tag{32}$$

其中，$\angle$计算复数的辐角。此外，似然函数通常还包含对总流量的约束（Sun等人，2024年）。前向模型的整体似然函数则通过汇总不同时间和望远镜组合下的这些数据产品来表示:

$$\begin{aligned}
\ell(\boldsymbol{y} \mid \boldsymbol{x}) = &\underbrace{\sum_{t,\text{c}} \frac{\left\|\mathcal{H}^{\text{cph}}_{t,\text{c}}(\boldsymbol{x}) - \boldsymbol{y}^{\text{cph}}_{t,\text{c}}\right\|_2^2}{2\beta^2_{\text{cph}}}}_{\chi^2_{\text{cph}}} + \underbrace{\sum_{t,\text{c}} \frac{\left\|\mathcal{H}^{\text{camp}}_{t,\text{c}}(\boldsymbol{x}) - \boldsymbol{y}^{\text{camp}}_{t,\text{c}}\right\|_2^2}{2\beta^2_{\text{camp}}}}_{\chi^2_{\text{camp}}} \\
&+ \rho\frac{\left\|\sum_i \boldsymbol{x}_i - \boldsymbol{y}^{\text{flux}}\right\|_2^2}{2}
\end{aligned} \tag{33}$$

其中，β_{cph} 和 β_{camp} 假定为已知标准差，$\boldsymbol{y}^{\text{flux}}$ 为精确测量值，而 $\rho > 0$ 为一个参数。前两项通常称为 χ^2 误差。我们使用Zheng 等人（2025b）提供的数据集和扩散模型的检查点进行实验，其中包含100张测试图像。我们还考虑了三种观测时间比例设置：100%、10%和3%。

流体数据同化。 纳维–斯托克斯方程是流体力学中的经典基准问题（Chandler 和 Kerswell，2013年）。它的应用范围广泛，从海洋动力学到气候建模；通常需要通过同化观测数据来校准数值预报的初始条件。

$$\begin{aligned}
\partial_t \boldsymbol{w}(\boldsymbol{x},t) + \boldsymbol{u}(\boldsymbol{x},t)\cdot\nabla\boldsymbol{w}(\boldsymbol{x},t) &= \nu\Delta\boldsymbol{w}(\boldsymbol{x},t) + f(\boldsymbol{x}), \\
\nabla\cdot\boldsymbol{u}(\boldsymbol{x},t) &= \mathbf{0}, \\
\boldsymbol{w}(\boldsymbol{x},0) &= \boldsymbol{w}_0(\boldsymbol{x}), \\
\boldsymbol{x} \in (0,2\pi)^2, t &\in (0,T]
\end{aligned} \tag{34}$$

其中，$\boldsymbol{u}$ 是速度场，$\boldsymbol{w} = \nabla \times \boldsymbol{u}$ 是涡量，$\boldsymbol{w}_0$ 是初始涡量，$\nu \in \mathbb{R}_+$是运动黏度系数，$f(\boldsymbol{x})$是强迫函数。我们考虑一个逆问题：根据$T = 1$时刻最终状态的稀疏含噪测量，恢复初始涡量场$\boldsymbol{w}_0$。观测模型如下:

$$\boldsymbol{y} = S\mathcal{F}(\boldsymbol{w}_0) + \boldsymbol{\eta} \tag{35}$$

其中，S 是采样算子，$\boldsymbol{\eta}$ 是测量噪声，$\mathcal{F}(\cdot)$是数值求解器的正向过程。我们使用伪谱方法（Li 等人 2020）求解纳维–斯托克斯方程。由于通过此类求解器进行反向传播存在固有困难且数值不稳定，我们将前向模型视为黑盒函数。我们的目标是根据最终时刻 $T = 1$、尺寸为 128×128 的稀疏含噪测量，恢复二维初始涡量场。我们使用 Zheng 等人（2025b）提供的数据集进行测试，其中包含 10 个不同的初始涡量场。我们考虑三种设置来模拟稀疏测量，包括 2×、 4×和 8× 下采样。所有测量均包含噪声，噪声水平为 $\sigma_y = 2.0$。

B2：实现细节

计算硬件和软件： 所有实验，包括计算时间和内存使用量测量，均在单块 NVIDIA A100 GPU 上进行。软件环境由 Python 3.9 和 PyTorch 2.4 组成。

基于梯度的方法的实现。 对于基于梯度的扩散逆问题求解器，包括 DDRM（Kawar 等人 2022a）、DPS（Chung 等人 2022）、RED-diff（Chung 等人 2022）和 DAPS（Zhang 等人 2025），我们使用各自作者提供的官方代码库。由于没有 ΠGDM（Song 等人 2023）的官方实现，我们根据原论文提供的伪代码忠实地复现了该方法。除非另有说明，我们均遵循这些方法的原始超参数配置。

无梯度方法的实现。对于我们提出的 CPS，所有任务均采用一致的超参数

方法	扩散 调度	时间步	引导 尺度	数量 更新次数	数量 粒子数	数量 前向调用次数
SCG（Huang 等人，2024）	VP（Song 等人，2020年）	1000	-	-	64	64k
DPG（Tang 等人，2024年）	VP（Song 等人，2020年）	1000	180.0	1	64	64k
EnKG（Zheng 等人，2025a）	EDM（Karras、Laine 和 Aila，2019年）	100	2.0	2	320	64k
CPS（本文）	VP（Song 等人，2020年）	500	-	-	64	64k

表4：用于基于扩散的逆问题求解的所评估无梯度方法的超参数配置。不适用于特定方法的超参数以短横线（–）表示。为进行公平比较，我们校准了每种方法的粒子数量，以确保前向观测模型调用的总次数保持一致。

设置。我们采用 500 步的随机 DDIM 求解器（Song、Meng 和 Ermon，2020年），并在每一步维护一个由 64 个粒子组成的集合。为提高鲁棒性，我们采用了重启策略。根据我们的观察，重启策略在生成过程后期的效果会减弱，因此仅在前 20% 的采样步骤中应用该策略。我们将重启次数 N_r 设为 5，这提升了解的质量，同时没有显著增加计算负担。为了与其他无梯度基线方法（SCG、DPG、EnKG）进行公平比较，我们调整了这些方法的粒子数，使其与 CPS 使用的前向模型调用总数一致。正文中的消融研究表明，增加粒子数通常能提升这些方法的性能。为保持一致，我们遵循这些方法官方实现中的默认超参数设置，粒子数除外。具体参数设置汇总于表4。

我们评估中所使用的对比方法的官方代码仓库如下所列：

- DDRM（Kawar 等人 2022a）：https://github.com/bahjat-kawar/ddrm
- DPS（Chung 等人 2022）：https://github.com/DPS2022/ diffusion-posterior-sampling
- RED-diff（Mardani 等，2023）：https://github.com/NVlabs/RED-diff
- DAPS（Zhang 等人 2025）：https://github.com/zhangbingliang2019/DAPS
- SCG（Huang 等人，2024年）：https://github.com/yjhuangcd/ rule-guided-music
- DPG（Tang 等人，2024年）：https://github.com/loveisbasa/ DPG
- EnKG（Zheng 等人，2025a）：https://github.com/devzhk/ enkg-pytorch

模型检查点和数据集。我们在基准数据集 FFHQ（Karras、Laine 和 Aila，2019年）上进行了实验。我们使用 Chung 等人（2022年）提供的预训练扩散模型，并从该数据集中随机抽取 100 张图像来计算所报告的指标。对于两个科学领域的逆问题——黑洞成像和流体数据同化——我们使用 InverseBench（Zheng 等人，2025b），这是一个专门的物理科学逆问题基准。InverseBench 提供了我们实验所需的预训练模型和测试数据。

附录 C：其他结果

图 9 展示了在 CPS 中采用重启策略的影响，其中使用不同的随机种子来初始化噪声。



图 9：使用不同随机种子初始化的重启策略消融研究。

附录 D：讨论与局限

我们的工作提出了约束粒子搜索（***CPS***），这是一种新的无梯度方法，旨在解决现有基于扩散模型的逆问题求解器所面临的性能问题。即使仅采样少量粒子，该方法也能取得良好效果。我们工作的一个核心理念贡献，是从被动候选选择转向主动粒子搜索。以 SCG 为代表的以往方法会在每一步丢弃大多数候选粒子，造成计算浪费。我们的实证研究结果表明，即使看似“糟糕”的粒子也包含有用信息，这有力地促使我们超越拒绝采样范式。通过将问题重新定义为一个受约束优化任务，CPS 能够有效整合所有候选粒子的信息，从而实现更高效、更稳健的采样过程。

尽管 CPS 在解决无梯度扩散逆问题方面取得了显著进展，但它并非没有局限，这也为未来研究提供了方向。CPS 仍需在每一步采样多个粒子，以估计代理模型。这会增加计算

成本，相较于无条件采样更高。一种潜在的方向是结合更快速的求解器（Xue 等人，2023年）或捷径采样（Liu 等人，2023年），并从初始猜测而非噪声开始，以进一步提高我们框架的效率。