arXiv:2603.01957v1 [math.OC] 2 Mar 2026

# 关于最小假设下 RED 算法收敛性及开放问题的札记

Yann Traonmilin* 和 J.-F. Aujol 波尔多大学、波尔多国立高等工程学院、法国国家科学研究中心、波尔多数学研究所，UMR 5251，F-33400，塔朗斯，法国 联系方式： yann.traonmilin@math.u-bordeaux.fr

摘要

本文针对测量值满足受限等距条件、所考虑的深度投影先验满足受限利普希茨条件的情形，给出一种改进的“去噪正则化”（RED）算法的收敛性结果。本研究引出了关于 RED 算法收敛性的开放问题。

## 1 引言

我们考虑如下观测模型

$$y = A\hat{x} + e \tag{1}$$

其中 $\hat{x} \in \Sigma \subset \mathbb{R}^N$、 $A \in \mathbb{R}^{m\times N}$ 和 $e \in \mathbb{R}^m$。集合 $\Sigma$ 是一个低维模型集。已有研究[11, 6] 表明，若 $A$ 满足受限等距性质（RIP），且给定的广义投影 $P_\Sigma$ 到 $\Sigma$满足受限利普希茨条件，则广义投影梯度下降（GPGD）迭代

$$x_{n+1} = P_\Sigma(x_n) - \mu A^T(AP_\Sigma(x_n) - y) \tag{2}$$

能够以线性速率稳定地恢复$\hat{x}$，即存在 $C > 0$ 使得

$$\|x_n - \hat{x}\|_2 \leq (\delta\beta)^n \|x_0 - \hat{x}\|_2 + C\|e\|_2. \tag{3}$$

这一结果已扩展至考虑模型误差和近似受限利普希茨投影的情形[5]。有人提出了这样一种模型，用于描述逆问题中即插即用（PnP）方法的收敛，其中 $P_\Sigma$ 由通用去噪器计算。事实证明，这种数学模型有助于识别驱动PnP可辨识性与收敛的关键几何参数。后续研究提出了深度投影先验的正则化方法，以增强这些性质 [6, 5]。据我们所知，RIP和受限利普希茨常数是目前已知能够导出此类收敛保证的最弱假设。该理论框架也已扩展至用于逆问题的扩散模型[7]。

另一个问题是，另一种流行的即插即用方案——通过去噪进行正则化（RED） [9, 8, 10, 2] ——在相同的最小假设下是否也能获得收敛保证。RED 迭代定义如下：

$$x_{n+1} = x_n - \mu A^T(Ax_n - y) - \lambda(x_n - P_\Sigma(x_n)). \tag{4}$$

本文证明，在类似（但更强）的假设下，经过轻微修改的 RED算法版本能够收敛，并提出由这些结果引出的一些开放问题。

## 2 符号说明

我们采用 [11] 中的形式化框架。

定义 2.1. 算子 $B$ 在割线集$\Sigma - \Sigma = \{x_1 - x_2$上具有受限等距常数 $\delta \in [0, 1)$: $x_1, x_2 \in \Sigma\}$ 当且仅当对所有 $x_1, x_2 \in \Sigma$,

$$\|(I - B)(x_1 - x_2)\|_2 \leq \delta \|x_1 - x_2\|_2. \tag{5}$$

我们用 $\delta_\Sigma(B)$ 表示 $B$ 的最小受限等距常数（RIC）。

定义 2.2（广义投影）。设 $\Sigma \subset \mathbb{R}^N$。到 $\Sigma$的（集值）广义投影是一个（集值）函数 $P$，满足对任意 $z \in \mathbb{R}^N$, $P(z) \subset \Sigma$。

为便于阅读，我们滥用符号：对于任意 $w \in P(z)$ 均成立的等式，记作 $P(z)$。我们在正交投影存在的集合上引入正交投影（即 $\ell^2$ 范数下的度量投影）。

定义 2.3（邻近集与正交投影）。设 $\Sigma \subset \mathbb{R}^N$。若对所有 $z \in \mathbb{R}^N$，均有

$$\left( \arg\min_{x \in \Sigma} \|x - z\|_2 \right) \neq \emptyset. \tag{6}$$

现在假设 $\Sigma$ 是一个邻近集，我们将到 $\Sigma$ 的正交投影定义为

$$P_\Sigma^\perp(z) := \arg\min_{x \in \Sigma} \|x - z\|_2. \tag{7}$$

注意， $P_\Sigma^\perp(z)$ 可能是一个集合值映射。

定义 2.4（受限利普希茨性质）。考虑广义投影 $P$。若对所有 $z \in \mathbb{R}^N$、$x \in \Sigma$、$u \in P(z)$，均有，则称 $P$ 相对于 $\Sigma$ 具有受限 $\beta$-利普希茨性质

$$\|u - x\|_2 \leq \beta \|z - x\|_2. \tag{8}$$

我们用 $\beta_\Sigma(P)$ 表示使 $P$ 具有受限 $\beta$-利普希茨性质的最小 $\beta$。

## 3 修改后的 RED 算法的收敛

对于参数 $\lambda \in [0, 1]$，我们指出，RED 可以解释为一个两步算法:

- 在当前迭代点 $x_n$ 与其到 $\Sigma$ 上的投影之间选取一点

$$z_n = (1 - \lambda)x_n + \lambda P_\Sigma(x_n) \in [x_n, P_\Sigma(x_n)] \tag{9}$$

  即 $x_n$ 与 $P_\Sigma(x_n)$ 的凸组合。

- 应用一个梯度下降步，其梯度（一个 $\ell^2$ 数据拟合项的梯度）在 $x_n$中计算:

$$x_{n+1} = z_n - \mu A^T(Ax_n - y). \tag{10}$$

通过改写 $z_n = x_n - \lambda(x_n - P_\Sigma(x_n))$，我们验证其对应于RED 迭代（4）。

我们建议考察如下修改后的 RED 算法：保持$z_n$ 的定义不变，但在 $z_n$ 而非 $x_n$中计算梯度步：

$$\begin{aligned} z_n &= (1-\lambda)x_n + \lambda P_\Sigma(x_n) \in [x_n, P_\Sigma(x_n)] \\ x_{n+1} &= z_n - \mu A^T(Az_n - y). \end{aligned} \tag{11}$$

我们考虑能够逼近到 $\Sigma$上的正交投影的情形（已证明其至少是受限 2-Lipschitz）[11]。

定理 3.1. 假设 $\Sigma$ 是一个邻近集。考虑修改后的 RED 迭代(11)，其中$\|P_\Sigma - P_\Sigma^\perp\|_2 \le \eta$。令 $\beta = \beta_\Sigma(P_\Sigma)$ 且 $\delta = \delta_\Sigma(\mu A^T A)$。假设 $r = (\delta\beta + |1-\lambda|\|I - \mu A^T A\|_{\text{op}}) < 1$，则

$$\|x_n - \hat{x}\|_2 \le r^n \|x_0 - \hat{x}\|_2 + \frac{1}{1-r}\left(|1-\lambda|\|I - \mu A^T A\|_{\text{op}}\eta + \mu\|A^T e\|_2\right). \tag{12}$$

证明。利用三角不等式，我们有

$$\begin{aligned} &\|x_{n+1} - \hat{x}\|_2 \\ &= \|z_n - \mu A^T(Az_n - y) - \hat{x}\|_2 \\ &= \|(I - \mu A^T A)(z_n - \hat{x}) + \mu A^T e\|_2 \\ &\le \|(I - \mu A^T A)(P_\Sigma(x_n) - \hat{x})) + (I - \mu A^T A)(z_n - P_\Sigma(x_n))\|_2 + \mu\|A^T e\|_2 \\ &= \|(I - \mu A^T A)(P_\Sigma(x_n) - \hat{x})) + (I - \mu A^T A)((1-\lambda)x_n + \lambda P_\Sigma(x_n) - P_\Sigma(x_n))\|_2 + \mu\|A^T e\|_2 \\ &= \|(I - \mu A^T A)(P_\Sigma(x_n) - \hat{x})) + (1-\lambda)(I - \mu A^T A)(x_n - P_\Sigma(x_n))\|_2 + \mu\|A^T e\|_2 \\ &\le \|(I - \mu A^T A)(P_\Sigma(x_n) - \hat{x}))\|_2 + \|(1-\lambda)(I - \mu A^T A)(x_n - P_\Sigma(x_n))\|_2 + \mu\|A^T e\|_2 \\ &\le \|(I - \mu A^T A)(P_\Sigma(x_n) - \hat{x}))\|_2 + |1-\lambda|\|I - \mu A^T A\|_{\text{op}}\|x_n - P_\Sigma(x_n)\|_2 + \mu\|A^T e\|_2. \end{aligned} \tag{13}$$

根据正交投影的定义，我们有 $\|x_n - P_\Sigma^\perp(x_n)\|_2 \le \|x_n - \hat{x}\|_2$。结合 RIC 与受限利普希茨性质，我们推出

$$\begin{aligned} &\|x_{n+1} - \hat{x}\|_2 \\ &\le \delta\beta\|x_n - \hat{x}\|_2 + |1-\lambda|\|I - \mu A^T A\|_{\text{op}}\|P_\Sigma^\perp(x_n) - \hat{x}\|_2 + |1-\lambda|\|I - \mu A^T A\|_{\text{op}}\eta + \mu\|A^T e\|_2 \\ &\le \delta\beta\|x_n - \hat{x}\|_2 + |1-\lambda|\|I - \mu A^T A\|_{\text{op}}\|x_n - \hat{x}\|_2 + |1-\lambda|\|I - \mu A^T A\|_{\text{op}}\eta + \mu\|A^T e\|_2 \\ &= (\delta\beta + |1-\lambda|\|I - \mu A^T A\|_{\text{op}})\|x_n - \hat{x}\|_2 + |1-\lambda|\|I - \mu A^T A\|_{\text{op}}\eta + \mu\|A^T e\|_2. \end{aligned} \tag{14}$$

令 $r = \delta\beta + |1-\lambda|\|I - \mu A^T A\|_{\text{op}}$，通过归纳法，与 [6]，类似，我们有

$$\|x_n - \hat{x}\|_2 \le r^n \|x_0 - \hat{x}\|_2 + \frac{1}{1-r}\left(|1-\lambda|\|I - \mu A^T A\|_{\text{op}}\eta + \mu\|A^T e\|_2\right). \tag{15}$$

□

我们可以看到，如果选择 $\lambda = 1$，就会得到 GPGD 迭代，并获得与修改后的 RED 相同的收敛速率（在投影满足额外正交性假设的情况下）。其他可行的 $\lambda$ 会导致收敛更慢（且可辨识性更低）。这留下了一个问题：在没有正交性假设的情况下（如 GPGD 中那样），采用一般的受限 Lipschitz 投影时，是否会收敛到稳定解。

# 4 经典RED的收敛性

在凸最小化的背景下，$x_n - P_\Sigma(x_n)$ 可以解释为某个凸函数的梯度（或者 $\Sigma$ 是一个凸集），[3]，此类算法的收敛性可以得到证明（通常具有次线性收敛率）。在非凸情形下，若 $x_n - P_\Sigma(x_n)$ 是全局Lipschitz的（且是某个函数的梯度），[4]则可以证明其收敛到临界点。为了在我们的经典RED情形下使用经典证明技巧证明收敛性，我们考虑 $\theta$ 使得 $0 < \lambda + \theta \leq 1$ 并将迭代（4）写成如下形式。

$$\begin{aligned} x_{n+1} &= (1-\lambda)x_n + \lambda P_\Sigma(x_n) - \mu A^T A(x_n - \hat{x}) \\ &= (1-\lambda-\theta)x_n + \lambda P_\Sigma(x_n) + \theta x_n - \mu A^T A(x_n - \hat{x})) \\ &= (1-\lambda-\theta)x_n + (\lambda+\theta)T(x_n) \end{aligned} \tag{16}$$

其中 $T$: $u \mapsto \frac{1}{\lambda+\theta}(\lambda P_\Sigma(u) + \theta u - \mu A^T A(u - \hat{x}))$。算子 $T_{\lambda+\theta}$: $x_n \mapsto (1-\lambda-\theta)x_n + (\lambda+\theta)T(x_n)$ 称为 $T$ 的平均算子。证明此类不动点迭代方案收敛的主要假设是，$T$ 是准非扩张的，即相对于 fix($T$) [1] 满足受限1-Lipschitz条件。事实上，我们验证了 $\hat{x} \subset$ fix($T$)。

应用于我们的情形时，我们需要相应的不等式

$$\begin{aligned} \|T(u) - \hat{x}\|_2 &\leq \|u - \hat{x}\|_2 \\ L := \|\frac{1}{\lambda+\theta}(\lambda P_\Sigma(u) + \theta u - \mu A^T A(u - \hat{x})) - \hat{x}\|_2^2 &\leq \|u - \hat{x}\|_2^2. \end{aligned} \tag{17}$$

我们可以写成

$$\begin{aligned} L &= \|(I - \frac{\mu}{\lambda+\theta}A^T A)(P_\Sigma(u) - \hat{x}) + \frac{1}{\lambda+\theta}(-\theta P_\Sigma(u) + \theta u - \mu A^T A(u - P_\Sigma(u))\|_2^2 \\ &= \|(I - \frac{\mu}{\lambda+\theta}A^T A)(P_\Sigma(u) - \hat{x}) + \frac{1}{\lambda+\theta}(\theta I - \mu A^T A)(u - P_\Sigma(u))\|_2^2 \\ &= \|w_1 + w_2\|_2^2 \end{aligned} \tag{18}$$

其中 $w_1 = (I - \frac{\mu}{\lambda+\theta}A^T A)(P_\Sigma(u) - \hat{x})$ 和 $w_2 = \frac{1}{\lambda+\theta}(\theta I - \mu A^T A)(u - P_\Sigma(u))$。

在条件允许时，通过对 $\frac{\mu}{\lambda+\theta}$ $\|w_1\|_2 \leq \delta$ $\frac{\mu}{\lambda+\theta}$ $\beta\|u - x\|_2 <$ 进行适当缩放，我们有 $\|u - x\|_2$ ( ) $\hat{}$ $\hat{}$。主要问题在于，在 RIP 条件下，我们通常无法用收缩项来界定，因为 $u - P_\Sigma(u) \notin \Sigma - \Sigma$。此外，如果模型集 $\Sigma$ 不具备凸性，那么在展开平方范数时，我们确实很难控制项 $\langle u - P_\Sigma(u)$ $P_\Sigma(u) - \hat{x}\rangle$ 的符号。 $w_2$

假设我们可以将 $u - P_\Sigma(u) = \sum_i \lambda_i v_i$ 分解为 $\sum_i \lambda_i = 1$ 和 $v_i \in \Sigma - \Sigma$（当 span$\Sigma = \mathbb{R}^N$ 和 $\Sigma$ 齐次时，这种分解是可能的）。利用 RIC 对 $w_2$ 进行界定的一种方法是使用这种分解。根据凸性，我们有

$$\|w_2\|_2^2 \leq \sum_i \lambda_i \left(\frac{\theta}{\lambda+\theta}\|(I - \frac{\mu}{\theta}A^T A)v_i\|_2\right)^2 \leq \left(\frac{\theta}{\lambda+\theta}\delta(\frac{\mu}{\theta}A^T A)\right)^2 \sum_i \lambda_i \|v_i\|_2^2 \tag{19}$$

对所有可能的分解 $(\lambda_i, v_i)_i, v_i \in \Sigma - \Sigma$ 取下确界，得到

$$\|w_2\|_2 \leq \frac{\theta}{\lambda+\theta}\delta(\frac{\mu}{\theta}A^TA)\|u - P_\Sigma(u)\|_{\Sigma-\Sigma} \tag{20}$$

其中 $\|z\|_{\Sigma-\Sigma} = \text{下确界}_{z_i \in \Sigma-\Sigma,\ \sum_i \lambda_i z_i = z,\ \sum_i \lambda_i = 1} \sqrt{\sum_i \lambda_i \|z_i\|_2^2}$ 是由 $\Sigma - \Sigma$ 诱导的原子范数（参见[12]）。现在，通常我们无法将距离 $u - P_\Sigma(u)$ 与 $u - \hat{x}$ 联系起来，除非 $P_\Sigma = P_\Sigma^\perp$ 是正交投影。在这种情况下，我们有 $\|u - P_\Sigma(u)\|_2$ 并且

$$\|w_2\|_2 \leq \frac{\theta}{\lambda+\theta}\delta(\frac{\mu}{\theta}A^TA)C_\Sigma\|u - \hat{x}\|_2 \tag{21}$$

其中 $C_\Sigma$ 是 $\|\cdot\|_2$ 与 $\|\cdot\|_{\Sigma-\Sigma}$ 之间的等价常数。因此，在满足以下条件时，可以实现拟非扩张性

$$\delta(\frac{\mu}{\lambda+\theta}A^TA)\beta + \frac{\theta}{\lambda+\theta}\delta(\frac{\mu}{\theta}A^TA)C_\Sigma \leq 1. \tag{22}$$

要实现这一点，例如，我们需要较小的 $\theta$、最优的 $\mu$ 和 $\lambda \approx 1$，以及 GPGD方案 $\delta(\mu A^TA)\beta < 1$ 在 $\delta(A^TA)\beta$ 较小时具有非常有利的可辨识性条件。常数 $C_\Sigma$ 或许会使这无法实现。这引出了以下开放问题：在这个理论框架下，是否有可能在更宽松的条件下，利用 RIC 和受限利普希茨条件证明经典RED迭代收敛?

鉴于上一节的结果，在何种理论框架下，才有理由使用（经典或修正的）RED 方法，而不是 GPGD——毕竟 RED 方法的收敛速率和可辨识性都不如 GPGD?

本文首次分析了在 RIC 和受限利普希茨条件下，修正版 RED算法的收敛性。我们希望这能成为证明原始 RED算法在类似假设下收敛的第一步。

# 参考文献

[1] H. H. Bauschke 和 P. L. Combettes. 希尔伯特空间中的凸分析与单调算子理论。[2] P. Cascarano、A. Benfenati、U. S. Kamilov 和 X. Xu. 基于去噪的约束正则化与自动参数选择。IEEE信号处理快报，31:556–560，2024年。[3] R. Cohen、M. Elad 和 P. Milanfar. 通过不动点投影进行去噪正则化（RED-Pro）。SIAM成像科学杂志，14(3):1374–1406，2021年。[4] S. Hurault、A. Leclaire 和 N. Papadakis. 用于实现收敛即插即用的梯度步去噪器。载于国际学习表征会议（ICLR'22），2022年。[5] A. Joundi、Y. Traonmilin 和 J.-F. Aujol. 从稀疏恢复到即插即用先验：通过广义投影梯度下降实现稳定恢复的权衡分析。arXiv预印本 arXiv:2512.07397，2025年。[6] A. Joundi、Y. Traonmilin 和 A. Newson. 深度投影先验的随机正交正则化。arXiv预印本 arXiv:2505.13078，2025年。

[7] O. Leong 和 Y. Traonmilin. 扩散先验的恢复理论：隐式先验算法的确定性分析。载于第29届人工智能与统计国际会议。[8] E. T. Reehorst 和 P. Schniter. 去噪正则化：澄清与新的诠释。IEEE计算成像汇刊，5(1):52–67，2018年。[9] Y. Romano、M. Elad 和 P. Milanfar. 小引擎，大能量：去噪正则化（RED）。SIAM成像科学杂志，10(4):1804–1844，2017年。[10] Y. Sun、J. Liu 和 U. Kamilov. 块坐标去噪正则化。神经信息处理系统进展，32，2019年。[11] Y. Traonmilin、J. F. Aujol 和 A. Guennec. 探索以线性收敛率恢复低维模型的最优算法。arXiv预印本 arXiv:2410.06607，2024年。[12] Y. Traonmilin 和 R. Gribonval. 希尔伯特空间中低维锥的稳定恢复：一个RIP统领一切。应用与计算调和分析，45(1):170–205，2018年。