Learning Pseudo-Contractive Denoisers for Inverse Problems

Deliang Wei [1] **Peng Chen** [1] **Fang Li** [1 2]

Abstract

Deep denoisers have shown excellent performance in solving inverse problems in signal and image processing. In order to guarantee the convergence, the denoiser needs to satisfy some Lipschitz conditions like non-expansiveness. However, enforcing such constraints inevitably compromises recovery performance. This paper introduces a novel training strategy that enforces a weaker constraint on the deep denoiser called pseudo-contractiveness. By studying the spectrum of the Jacobian matrix, relationships between different denoiser assumptions are revealed. Effective algorithms based on gradient descent and Ishikawa process are derived, and further assumptions of strict pseudo-contractiveness yield efficient algorithms using half-quadratic splitting and forward-backward splitting. The proposed algorithms theoretically converge strongly to a fixed point. A training strategy based on holomorphic transformation and functional calculi is proposed to enforce the pseudo-contractive denoiser assumption. Extensive experiments demonstrate superior performance of the pseudo-contractive denoiser compared to related denoisers. The proposed methods are competitive in terms of visual effects and quantitative values.

1. Introduction

Inverse problems aim to recover the potential signal from down sampled or corrupted obsevations. A typical inverse problem takes form of:

$$f = Ku + n, \tag{1}$$

where f is the observed signal, u is the potential signal, K is the degradation operator, and n is the noise following certain distributions. Different values of K and n correspond to different missions including denoising, deblurring, inpainting, super-resolution, and medical imaging. In order to recover u from f, a variational approach is considered:

$$\hat{u} = \arg\min_{u \in V} F(u) + G(u; f), \tag{2}$$

where V is the Hilbert space, F denotes the prior regularization term, and G is the data fidelity term. Typical choices for F include total variation (Rudin et al., 1992) and its extensions (Bredies et al., 2010), weighted nuclear norm (Gu et al., 2014), group-based low rank prior (Mairal J, 2009) et al.. First order methods are employed to solve (2), such as the alternating direction method with multipliers (ADMM) (Boyd et al., 2011):

$$\begin{aligned} u^{k+1} &= \mathrm{Prox}_{\frac{F}{\beta}}(v^k - b^k), \\ v^{k+1} &= \mathrm{Prox}_{\frac{G}{\beta}}(u^{k+1} + b^k), \\ b^{k+1} &= b^k + u^{k+1} - v^{k+1}, \end{aligned} \tag{3}$$

where $\beta > 0$. For a given proper, closed, and convex function $F : V \to (-\infty, \infty]$, the proximal operator $\mathrm{Prox}_F : V \to V$ is defined as:

$$\mathrm{Prox}_F(y) = \arg\min_{x \in V} F(x) + \frac{1}{2}\|x - y\|^2. \tag{4}$$

Noticing that $\mathrm{Prox}_{\frac{F}{\beta}}(\cdot)$ is a Gaussian denoiser, Venkatakrishnan et al. proposed to replace the u-subproblem in (3) with arbitrary Gaussian denoiser D_σ in a plug-and-play (PnP) fashion, that is $u^{k+1} = \mathrm{D}_\sigma(v^k - b^k)$, and arrived at PnP-ADMM algorithm (Venkatakrishnan et al., 2013). Here, D_σ is a Gaussian denoiser with denoising strength σ. Throughout the paper, we let $\beta = \frac{1}{\sigma^2}$ as suggested in (Chan et al., 2016; Zhang et al., 2017b; 2021).

Interestingly, PnP-ADMM, along with other PnP methods, has demonstrated remarkable recovery effects in a diverse range of areas, such as bright field electron tomography (Sreehari et al., 2016), camera image processing (Heide et al., 2014), low-dose CT imaging (Venkatakrishnan et al., 2013; Peng et al., 2023), image denoising (Le Pendu & Guillemot, 2023; Wei et al., 2023a;b), deblurring (Zhang et al., 2017b; Laroche et al., 2023a), inpainting (Zhu et al.,

[1]School of Mathematical Sciences, Key Laboratory of MEA(Ministry of Education) & Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China [2]Chongqing Key Laboratory of Precision Optics, Chongqing Institute of East China Normal University, Chongqing 401120, China. Correspondence to: Fang Li <fli@math.ecnu.edu.cn>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

2023), and super-resolution (Laroche et al., 2023b). However, the analysis of convergence is challenging due to the inherent black box nature of D_σ. Ensuring the convergence of PnP algorithms with weaker assumptions and more powerful denoisers has emerged as a demanding research topic, requiring further investigation and exploration.

The existing approaches to guarantee the convergence of PnP methods can be classified into two categories.

The first class aims to find a potential function $F : V \to (-\infty, \infty]$, such that $\mathrm{D}_\sigma = \nabla F$ or $\mathrm{D}_\sigma = \mathrm{Prox}_F$. By studying the Jacobian matrix $\mathrm{J}(x) = \nabla \mathrm{D}_\sigma(x)$, Sreehari et al. first proved that when J is symmetric with eigenvalues in $[0, 1]$ for any $x \in V$, there exists some proper, closed, and convex F, such that $\mathrm{D}_\sigma(\cdot) = \mathrm{Prox}_{\frac{F}{\beta}}(\cdot)$ is indeed a proximal operator (Sreehari et al., 2016). This assumption may be too strong, that most denoisers like non-local means (NLM) (Buades et al., 2005), BM3D (Dabov et al., 2006), DnCNN (Zhang et al., 2017a), and UNet (Ronneberger et al., 2015) violate it. Romano et al. proposed the regularization by denoising (RED) method, which is more flexible than PnP-ADMM (Romano et al., 2017). The RED prior term takes the form of $F(x) = \frac{1}{2}\langle x, x - \mathrm{D}_\sigma(x)\rangle$. Romano et al. proved that when D_σ is locally homogeneous, $\nabla \mathrm{D}_\sigma$ is symmetric with spectral radius less than one, one has $\nabla F(x) = x - \mathrm{D}_\sigma(x)$, and that PnP-GD and PnP-ADMM with RED prior converge. Yet the assumptions are impractical. As reported by Reehorst and Schniter, deep denoisers may not satisfy these assumptions (Reehorst & Schniter, 2018). In (Cohen et al., 2021), instead of training a Gaussian denoiser D_σ, Cohen et al. parameterized an implicit convex function $F : V \to (-\infty, \infty)$ with a neural network by enforcing non-negative weights and convex, non-decreasing activations, such that F is convex, and $\mathrm{D}_\sigma(\cdot) = \nabla F(\cdot) : V \to V$ outputs clean images. By doing so, an implicit convex prior F is obtained, and a convergent algorithm based on gradient decent (GD) is derived. Unfortunately, experimental results have demonstrated that employing a convex regularization term compromises the effectiveness of recovery. In (Hurault et al., 2022a), Hurault et al. proposed the gradient step (GS) denoiser $\mathrm{D}_\sigma = \mathrm{I} - \nabla F$, where F is parameterized by DRUNet (Zhang et al., 2021). In (Hurault et al., 2022b), Hurault et al. proposed the proximal DRUNet (Prox-DRUNet), which requires that ∇F is L-Lipschitz with $L \leq 0.5$. Under these assumptions, they proved the convergence of PnP with half-quadratic splitting (PnP-HQS) and PnP-ADMM. Nonetheless, the assumptions may still be too strong: the constraint on $L \leq 0.5$ has been proved to be too restrictive and hindered the denoising performance, see (Hurault et al., 2022b).

The second class of research investigates the assumptions of D_σ under which PnP has a fixed-point convergence. In the work of Sreehari et al. (Sreehari et al., 2016), the Jacobian matrix J of the denoiser D_σ is assumed to be symmetric, with eigenvalues lying inside $[0, 1]$. Then D_σ is firmly non-expansive. As a result, PnP-ADMM converges to a fixed point. Inspired by this pioneer work, Chan et al. analyzed convergence with a bounded denoiser assumption (Chan et al., 2016). The denoising strength decreases to ensure the convergence. In (Buzzard et al., 2018), Buzzard et al. explained PnP via the framework of consensus equilibrium. The convergence is proved for non-expansive denoisers. However, as reported in (Chan et al., 2016), deep denoisers are in general expansive. In (Sun et al., 2019), Sun et al. analyzed the convergence of PnP with proximal gradient descent (PnP-PGM) under the assumption that D_σ is θ-averaged ($\theta \in (0, 1)$). The averagedness assumption is too restricted, since many denoisers cannot be considered as averaged denoiser (Laumont et al., 2023). In (Tirer & Giryes, 2018), Tirer and Giryes assumed a bounded D_σ with contractive projections on a subspace, and provided an upper bound on the error of the recovered signal and the true signal. Nevertheless, the assumptions are difficult to validate. In (Ryu et al., 2019), Ryu et al. enforced the contractiveness of $\mathrm{I} - \mathrm{D}_\sigma$ by real spectral normalization (RealSN), which normalized the spectral norm of each layer. However, RealSN is time consuming, and is designed specifically for denoisers with cascade residual learning structures like DnCNN, and thus is not suitable for other networks like UNet. In (Pesquet et al., 2021), Pesquet proved the convergence of PnP-FBS when $2\,\mathrm{D}_\sigma - \mathrm{I}$ is non-expansive. Since that the non-expansiveness of D_σ can be drawn from the non-expansiveness of $2\,\mathrm{D}_\sigma - \mathrm{I}$, the constraint is more restrictive, and the performance of the denoiser is not satisfying.

Contributions. As discussed above, in order to guarantee the convergence of PnP and RED algorithms, the previous works assume the Lipschitz properties of the denoisers. However, enforcing such assumptions inevitably comprises the denoising performance. To address these issues, in this paper, we propose convergent plug-and-play methods with pseudo-contractive denoisers. Overall, our main contributions are threefold:

- The assumption regarding the denoiser is pseudo-contractiveness, which is weaker than that of existing methods. To ensure this assumption, an effective training strategy has been proposed based on holomorphic transformation and functional calculi.
- Convergent plug-and-play Ishikawa methods based on GD, HQS, and FBS are proposed. The global convergence results are established.
- Numerical experiments show that the proposed methods are competitive compared with other closely related methods in terms of visual effects, and quantitive values.

2. Pseudo-contractive Denoisers

Let V be the real Hilbert space with the inner product $\langle \cdot, \cdot \rangle$. A mapping $\mathrm{D} : V \to V$ is θ-averaged, $\theta \in (0, 1]$, if

$$\mathrm{D} = \theta\,\mathrm{N} + (1-\theta)\,\mathrm{I}, \tag{5}$$

where N is a non-expansive mapping, that is $\forall x, y \in V$, $\|\mathrm{N}(x) - \mathrm{N}(y)\| \leq \|x - y\|$, I is the identity mapping, $\|\cdot\|$ is the induced norm from $\langle \cdot, \cdot \rangle$. Some other closely-related Lipschitz assumptions are reviewed in Appendix A.

D is pseudo-contractive (Rafiq, 2007; Weng, 1991; Hicks & Kubicek, 1977), if there exists $0 \leq k \leq 1$, such that,

$$\begin{aligned} &\|\mathrm{D}(x) - \mathrm{D}(y)\|^2 \leq \|x - y\|^2 \\ &+k\|(\mathrm{I}-\mathrm{D})(x) - (\mathrm{I}-\mathrm{D})(y)\|^2,\ \forall x, y \in V. \end{aligned} \tag{6}$$

When $0 \leq k < 1$, D is strictly pseudo-contractive (Chidume, 1987; Weng, 1991). Non-expansiveness is a special case of pseudo-contractive with $k = 0$.

Lemma 2.1 gives an equivalent definition of k-strictly pseudo-contractive mappings, and therefore gives the relationship between strictly pseudo-contractive mappings and the averaged mappings in the form of (5).

Lemma 2.1. *(proof in Appendix C) The following two statements are equivalent:*
- $\mathrm{D} : V \to V$ *is k-strictly pseudo-contractive with $k < 1$;*
- $\mathrm{D} : V \to V$ *can be written as* $\mathrm{D} = \frac{1}{1-k}\mathrm{N} - \frac{k}{1-k}\mathrm{I}$, *where* N *is non-expansive.*

It is worth noting that D is pseudo-contractive, if and only if $\mathrm{I} - \mathrm{D}$ is monotone.

Lemma 2.2. *(proof in Appendix D) Let* $\mathrm{D} : V \to V$ *be a mapping in the real Hibert space V. Then,* D *is pseudo-contractive, if and only if* $\mathrm{I} - \mathrm{D}$ *is monotone, that is*

$$\langle (\mathrm{I}-\mathrm{D})(x) - (\mathrm{I}-\mathrm{D})(y), x - y \rangle \geq 0. \tag{7}$$

The relationships between these properties are:
Firmly Non-expansive $\Rightarrow$ Averaged $\Rightarrow$ Non-expansive $\Rightarrow$ Pseudo-contractive.

It has been reported in (Hurault et al., 2022b) that, imposing non-expansive denoiser alters its denoising performance. Pseudo-contractiveness enlarges the range of the denoisers in the following sense. Let D be a deep Gaussian denoiser, which inputs a noisy image and outputs a clean image. In this setting, $\mathrm{I} - \mathrm{D}$ outputs the predicted noise. Pseudo-contractive D means that the difference between two output clean images is smaller than the sum of the difference between the input noisy images and the difference between the predicted noises. As a result, Pseudo-contractiveness is a weaker assumption on the deep denoisers than non-expansiveness, averagedness, and firmly non-expansiveness.

We further explore the potential relationships between different assumptions on the denoisers by studying the spectrum distribution. Let $\mathrm{D} \in \mathcal{C}^1[V]$, and $\mathrm{J}(x) = \nabla_x \mathrm{D}$ be the Jacobian matrix at point $x \in V$ of D. By the mean value theorem, (7) can be rewritten as

$$\langle (\mathrm{I} - \mathrm{J}^{\mathrm{T}}(\xi))(x - y), x - y \rangle \geq 0, \xi = \xi(x, y) \in V. \tag{8}$$

Thus D is pseudo-contractive, if there holds

$$\langle (\mathrm{I} - \mathrm{J}^{\mathrm{T}})(x - y), x - y \rangle \geq 0, \tag{9}$$

for any $x, y, \xi \in V$, $\mathrm{J} = \mathrm{J}(\xi)$. We refer (9) to the *regularity condition* of pseudo-contractiveness. J can be decomposed into a symmetric part $\mathrm{S} = \frac{1}{2}(\mathrm{J} + \mathrm{J}^{\mathrm{T}})$ and an anti-symmetric part $\mathrm{A} = \frac{1}{2}(\mathrm{J} - \mathrm{J}^{\mathrm{T}})$. For any $x, y \in V$, we have

$$\langle (\mathrm{I} - \mathrm{J}^{\mathrm{T}})(x - y), x - y \rangle = \langle (\mathrm{I} - \mathrm{S})(x - y), x - y \rangle. \tag{10}$$

As a result, condition (9) is equivalent to that any eigenvalue of S is not larger than 1. Thus, the real part of any eigenvalue of J is smaller than 1. That is, the eigenvalue of J for pseudo-contractive D lies inside the half plane, $\mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : real(z) \leq 1\}$, where $\mathrm{Sp}(\mathrm{J})$ denotes the spectrum set of J:

$$\mathrm{Sp}(\mathrm{J}) := \bigcup_{x \in V} \mathrm{Sp}(\mathrm{J}(x)). \tag{11}$$

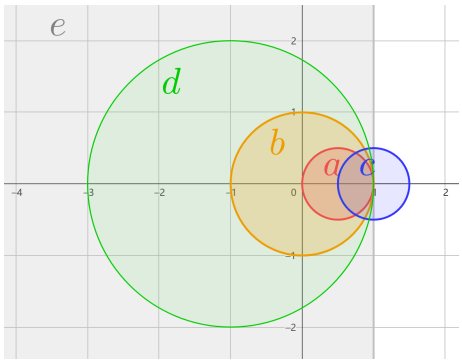


Figure 1. Spectrum distributions on the complex plane for the Jacobian under different assumptions. (a) Firmly non-expansiveness, specifically $\frac{1}{2}$-averagedness, $\mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : |2z - 1| \leq 1\}$; (b) Non-expansiveness, $\mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : |z| \leq 1\}$; (c) Contractiveness of $\mathrm{I} - \mathrm{J}$ with $r = \frac{1}{2}$, $\mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : |z - 1| \leq \frac{1}{2}\}$; (d) k-strictly pseudo-contractiveness with $k = \frac{1}{2}$, $\mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : |z + 1| \leq 2\}$; (e) Pseudo-contractiveness, $\mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : real(z) \leq 1\}$.

Similarly, we give the *regularity conditions* for the Jacobian J under θ-averaged, and pseudo-contractive assumptions on the denoiser D, as well as the distribution regions $\mathrm{Sp}(\mathrm{J})$ in (12)-(14). $\|\cdot\|_*$ denotes the spectral norm. Conditions for non-expansive, firmly non-expansive, and contractive D is given in Appendix A. Note that these regularity conditions

are sufficient conditions for a denoiser to satisfy the corresponding assumptions.

• θ-averaged D ($\theta \in (0,1]$):

$$\begin{aligned} &\left\| \left[\left(1-\tfrac{1}{\theta}\right) \mathrm{I} + \tfrac{1}{\theta} \mathrm{J} \right] \right\|_* \leq 1, \\ &\mathrm{Sp}(\mathrm{J}) \subset \left\{ z \in \mathbb{C} : \left| 1 - \tfrac{1}{\theta} + \tfrac{1}{\theta} z \right| \leq 1 \right\}. \end{aligned} \tag{12}$$

• k-strictly pseudo-contractive D ($k < 1$):

$$\begin{aligned} &\| k\,\mathrm{I} + (1-k)\,\mathrm{J} \|_* \leq 1, \\ &\mathrm{Sp}(\mathrm{J}) \subset \{ z \in \mathbb{C} : |(1-k)z + k| \leq 1 \}. \end{aligned} \tag{13}$$

• Pseudo-contractive D:

$$\begin{aligned} &\langle (\mathrm{I} - \mathrm{J}^{\mathrm{T}})(x-y), x-y \rangle \geq 0, \\ &\mathrm{Sp}(\mathrm{J}) \subset \{ z \in \mathbb{C} : real(z) \leq 1 \}. \end{aligned} \tag{14}$$

3. The proposed algorithm

In this section, we propose PnP-based algorithms to solve (2) using (strictly) pseudo-contractive denoisers. Before that, we briefly review three existing PnP methods.

Gradient descent (GD) method solves (2) by

$$u^{n+1} = u^n - \alpha(\nabla F + \nabla G)u^n, \tag{15}$$

where $\alpha > 0$ is the step size. When F is parameterized by a neural network D_σ, ∇F is often replaced by $\mathrm{I} - \mathrm{D}_\sigma$ (Romano et al., 2017). Then, PnP-GD takes the form of

$$u^{n+1} = [(1-\alpha)\,\mathrm{I} + \alpha\,\mathrm{T}]u^n, \mathrm{T} = \mathrm{D}_\sigma - \nabla G. \tag{16}$$

Unlike PnP-GD, many PnP methods can be written as the composition of two mappings. For example, the iterations of PnP-HQS and PnP-FBS to solve (2) takes the form of

$$\begin{aligned} &\text{PnP-HQS:} \quad u^{n+1} = \mathrm{T}(u^n),\ \mathrm{T} = \mathrm{D}_\sigma \circ \mathrm{Prox}_{\frac{G}{\beta}}, \\ &\text{PnP-FBS:} \quad u^{n+1} = \mathrm{T}(u^n),\ \mathrm{T} = \mathrm{D}_\sigma \circ (\mathrm{I} - \lambda \nabla G). \end{aligned} \tag{17}$$

T in PnP-GD is the sum of D_σ and $-\nabla G$, while T in PnP-HQS and PnP-FBS is composed of two mappings. When D_σ is assumed to be pseudo-contractive, it is necessary to study the property of T in these cases. Lemma 3.1 gives the condition that the sum of D_σ and $-\nabla G$ is pseudo-contractive.

Lemma 3.1. *(proof in Appendix E) Let* D *be pseudo-contractive, and* G *be proper, closed, and convex. Then* $\mathrm{T} = \mathrm{D} - \nabla G$ *is also pseudo-contractive.*

Lemma 3.2 gives the condition that a strictly pseudo-contractive mapping composed with an averaged mapping is still pseudo-contractive.

Lemma 3.2. *(proof in Appendix F) Let* D *be* k*-strictly pseudo-contractive, and* P *be* θ*-averaged,* $k, \theta \in (0,1]$*. If* $k < 1 - \theta$*, the composite operator* $\mathrm{D} \circ \mathrm{P}$ *is* l*-strictly pseudo-contractive, where*

$$0 \leq l = \frac{k(1-\theta)}{(1-\theta) - k\theta} < 1. \tag{18}$$

If $k = 1 - \theta$*,* $\mathrm{D} \circ \mathrm{P}$ *is pseudo-contractive. Besides, when* $k < 1$*,* $\mathrm{D} \circ \mathrm{P}$ *is* $\frac{1+k}{1-k}$*-Lipschitz.*

By Lemmas 3.1-3.2, when D_σ is (strictly) pseudo-contractive, T is pseudo-contractive. We need a special iteration schemes to find the fixed point of a pseudo-contractive mapping T.

Ishikawa proposed the following process to find the fixed point of a Lipschitz pseudo-contractive mapping T over a compact convex set K (Ishikawa, 1974). He proved the following theorem.

Theorem 3.3. *Let* K *be a compact convex subset of a Hilbert space* V*,* $\mathrm{T} : K \to K$ *is a Lipschitz and pseudo-contractive mapping, and* $x^0 \in K$*, then the sequence* $\{x^n\}$ *converges strongly to a fixed point of* T*, where* x^n *is defined iteratively for* $n \geq 0$ *by:*

$$\begin{aligned} y^n &= (1-\beta_n)x^n + \beta_n\,\mathrm{T}\,x^n, \\ x^{n+1} &= (1-\alpha_n)x^n + \alpha_n\,\mathrm{T}\,y^n, \end{aligned} \tag{19}$$

where α^n, β^n *satisfy*

$$0 \leq \alpha_n \leq \beta_n < 1,\ \lim_{n\to\infty} \beta_n = 0,\ \sum_n \alpha_n \beta_n = \infty. \tag{20}$$

Now we extend the existing PnP-GD, PnP-HQS, and PnP-FBS to the Ishikawa process. According to PnP-GD in (16), by letting $\mathrm{T} = \mathrm{D}_\sigma - \nabla G$, we propose PnPI-GD, an abbreviation for PnP Ishikawa gradient descent in Algorithm 1. Theorem 3.4 gives the global convergence of PnPI-GD.

Theorem 3.4. *(proof in Appendix G)* K *is a compact convex set in* V*. Let* $\mathrm{D}_\sigma : K \to K$ *be Lipschitz pseudo-contractive,* $G : K \to K$ *be differentiable, proper, closed, and convex, with Lipschitz gradient* ∇G*.* $\{\alpha_n\}, \{\beta_n\}$ *be two sequences satisfying (20). Assume that* $Fix(\mathrm{D}_\sigma - \nabla G) \neq \emptyset$*. Then,* u^n *generated by PnPI-GD in Algorithm 1 converges strongly to a fixed point in* $Fix(\mathrm{D}_\sigma - \nabla G)$*.*

Algorithm 1 PnPI-GD

Given $\mathrm{D}_\sigma, \{\alpha_n\}, \{\beta_n\}, u^0, N$.
for $n = 0 : N-1$ **do**
  $v^n = (1-\beta_n)u^n + \beta_n(\mathrm{D}_\sigma(u^n) - \nabla G(u^n))$
  $u^{n+1} = (1-\alpha_n)u^n + \alpha_n(\mathrm{D}_\sigma(v^n) - \nabla G(v^n))$
end for
Return u^N

If T takes the form of PnP-HQS as in (17), $\mathrm{T} = \mathrm{D}_\sigma \circ \mathrm{Prox}_{\frac{G}{\beta}}$, we arrive at PnPI-HQS in Algorithm 2. When

Algorithm 2 PnPI-HQS

Given $\mathrm{D}_\sigma, \{\alpha_n\}, \{\beta_n\}, u^0, N$
for $n = 0 : N-1$ **do**
  $x^n = \mathrm{Prox}_{\frac{G}{\beta}}(u^n)$
  $v^n = (1-\beta_n)u^n + \beta_n \mathrm{D}_\sigma(x^n)$
  $y^n = \mathrm{Prox}_{\frac{G}{\beta}}(v^n)$
  $u^{n+1} = (1-\alpha_n)u^n + \alpha_n \mathrm{D}_\sigma(y^n)$
end for
Return u^N

Algorithm 3 PnPI-FBS

Given $\mathrm{D}_\sigma, \{\alpha_n\}, \{\beta_n\}, \lambda, u^0, N$
for $n = 0 : N-1$ **do**
  $x^n = u^n - \lambda \nabla G(u^n)$
  $v^n = (1-\beta_n)u^n + \beta_n \mathrm{D}_\sigma(x^n)$
  $y^n = v^n - \lambda \nabla G(v^n)$
  $u^{n+1} = (1-\alpha_n)u^n + \alpha_n \mathrm{D}_\sigma(y^n)$
end for
Return u^N

T is the Forward-Backward Splitting (FBS) operator, $\mathrm{T} = \mathrm{D}_\sigma \circ (\mathrm{I} - \lambda \nabla G)$, we arrive at PnPI-FBS in Algorithm 3. The corresponding convergence results are given in Theorem 3.5 and Theorem 3.6, respectively.

Theorem 3.5. *(proof in Appendix H) K is a compact convex set in V. Let $\mathrm{D}_\sigma : K \to K$ be Lipschitz k-strictly pseudo-contractive, $G : K \to K$ be proper, closed, and convex, ∇G is γ-cocoercive. $\{\alpha_n\}, \{\beta_n\}$ be two sequences satisfying (20). Assume that $Fix(\mathrm{D}_\sigma \circ \mathrm{Prox}_{\frac{G}{\beta}}) \neq \emptyset$. Then, u^n generated by PnPI-HQS in Algorithm 2 converges strongly to a fixed point in $Fix(\mathrm{D}_\sigma \circ \mathrm{Prox}_{\frac{G}{\beta}})$, if $k \leq \frac{2\gamma+1}{2\gamma+2}$.*

Theorem 3.6. *(proof in Appendix I) K is a compact convex set in V. Let $\mathrm{D}_\sigma : K \to K$ be Lipschitz k-strictly pseudo-contractive, $G : K \to K$ be proper, closed, and convex, ∇G is γ-cocoercive. $\{\alpha_n\}, \{\beta_n\}$ be two sequences satisfying (20). Assume that $Fix(\mathrm{D}_\sigma \circ (\mathrm{I} - \lambda \nabla G)) \neq \emptyset$. Then, u^n generated by PnPI-FBS in Algorithm 3 converges strongly to a fixed point in $Fix(\mathrm{D}_\sigma \circ (\mathrm{I} - \lambda \nabla G))$, if $0 \leq \lambda \leq 2\gamma$, and $k \leq 1 - \frac{\lambda}{2\gamma}$.*

Remark 3.7. For a proper, closed, convex, and differentiable G, ∇G is 0-cocoercive. As a result, according to Theorem 3.5, PnPI-HQS converges, if $k \leq \frac{1}{2}$. Similarly, when we select $\lambda \in [0, \gamma]$ in PnPI-FBS, we have $k = \frac{1}{2} \leq 1 - \frac{\lambda}{2\gamma}$. That is, a $\frac{1}{2}$-strictly pseudo-contractive denoiser D_σ satisfies the conditions in Theorems 3.5-3.6.

4. Training strategy

In this section, we propose an effective training strategy to ensure that the denoiser is pseudo-contractive. Let p be the distribution of the training set of clean images, and $[\sigma_{\min}, \sigma_{\max}]$ be the interval of the noise level. Inspired by (Pesquet et al., 2021), we utilize the spectral regularization technique to ensure the pseudo-contractive assumption by regularizing the Jacobian.

Strictly pseudo-contractiveness. In order to ensure the denoisers to be k-strictly pseudo-contractive, we need $\|k\,\mathrm{I} + (1-k)\,\mathrm{J}\|_* \leq 1$. Let θ be the parameters of the denoiser D_σ. An optimal $\hat{\theta}$ is a solution to the following problem:

$$\mathbb{E}\|\,\mathrm{D}_\sigma(x+\xi_\sigma;\theta) - x\|_1 + r\max\{\|k\,\mathrm{I} + (1-k)\,\mathrm{J}\,\|_*, 1-\epsilon\}, \tag{21}$$

where $x \sim p$, $\sigma \sim U[\sigma_{\min}, \sigma_{\max}], \xi_\sigma \sim \mathcal{N}(0, \sigma^2)$. The first term ensures that D_σ is a Gaussian denoiser, while the second term is the spectral regularization term. $r > 0$ is the balancing parameter, and $\epsilon \in (0, 1)$ is a parameter that controls the constraint.

Algorithm 4 Power iterative method

Given q^0 with $\|q^0\| = 1, \mathrm{J}, N$
for $n = 1 : N$ **do**
  $z^n = \mathrm{J}\, q^{n-1}$
  $q^n = \frac{z^n}{\|z^n\|}$
end for
Return $\lambda^N = (q^N)^{\mathrm{T}}\, \mathrm{J}\, q^N$

By the power iterative method (Golub & Van Loan, 2013), we can compute the spectral norm of J by Algorithm 4. The AutoGrad toolbox in Pytorch (Paszke et al., 2017) allows the calculation for $\mathrm{J}\,x$ and $\mathrm{J}^{\mathrm{T}}\,x$ with any vector x. Thus, z^n and λ^N can be obtained efficiently.

Pseudo-contractiveness. In order to train a pseudo-contractive denoiser, we need to constrain

$$\langle (\mathrm{I} - \mathrm{S})(x-y), x-y \rangle \geq 0, \forall x, y \in V. \tag{22}$$

Since S is symmetric, we can do functional calculi on S. We wish to find a holomorphic function $f : \mathbb{C} \to \mathbb{C}$ defined on the neighborhood of $\mathrm{Sp}(\mathrm{S})$, such that $f(\mathrm{S})$ is defined, and

$$f(\{z \in \mathbb{C} : real(z) \leq 1\}) = \{z \in \mathbb{C} : |z| \leq 1\}. \tag{23}$$

Then, by the spectral mapping theorem (Harte, 1972; Haase, 2005), there holds $\mathrm{Sp}(f(\mathrm{S})) = f(\mathrm{Sp}(\mathrm{S}))$. We choose the following function $f(z) = \frac{z}{z-2}, \forall z \neq 2$. Then f is holomorphic on the neighborhood of the spectrum set of a pseudo-contractive denoiser. Besides, f maps the half plane $\{z \in \mathbb{C} : real(z) \leq 1\}$ to the unit disk.

To ensure $\mathrm{Sp}(\mathrm{S}) \subset \{z \in \mathbb{C} : real(z) \leq 1\}$, we only need to constrain $\mathrm{Sp}(f(\mathrm{S})) \subset \{z \in \mathbb{C} : |z| \leq 1\}$. Note that S is symmetric, $\rho(\mathrm{S}) = \|\,\mathrm{S}\,\|_*$. As a result, we only need to

constrain $\|f(\mathrm{S})\|_* \leq 1$, because

$$\begin{aligned} & \|f(\mathrm{S})\|_* \leq 1 \Leftrightarrow \rho(f(\mathrm{S})) \leq 1 \\ \Leftrightarrow \quad & f(\mathrm{Sp}(\mathrm{S})) = \mathrm{Sp}(f(\mathrm{S})) \subset \{z \in \mathbb{C} : |z| \leq 1\} \\ \Leftrightarrow \quad & \mathrm{Sp}(\mathrm{S}) \subset f^{-1}(\{z \in \mathbb{C} : |z| \leq 1\}) \\ = \quad & \{z \in \mathbb{C} : real(z) \leq 1\} \\ \Rightarrow \quad & \text{The regularity condition (9) holds.} \\ \Rightarrow \quad & \mathrm{D}_\sigma \text{ is pseudo-contractive.} \end{aligned} \tag{24}$$

The above derivations can be summarized as the following theorem.

Theorem 4.1. *Let* $\mathrm{J} = \mathrm{J}(x) = \nabla_x \mathrm{D}_\sigma$ *be the differential of* $\mathrm{D}_\sigma : V \to V$ *at* $x \in V$. *Let* S *be the symmetric part of* J. $f(z) = \frac{z}{z-2}$ *is a holomorphic function on the neighborhood of* $\{z \in \mathbb{C} : real(z) \leq 1\}$. *Then* D_σ *is pseudo-contractive, if for any* $x \in V, \mathrm{S} = \mathrm{S}(x)$, *there holds* $\|f(\mathrm{S})\|_* \leq 1$, *where* $\|\cdot\|_*$ *denotes the spectral norm.*

The loss function for a pseudo-contractive denoiser is

$$\mathbb{E}\|\mathrm{D}_\sigma(x+\xi_\sigma;\theta) - x\|_1 + r\max\{\|(\mathrm{S}-2\,\mathrm{I})^{-1}\,\mathrm{S}\,\|_*, 1-\epsilon\}. \tag{25}$$

According to the power iterative method in Algorithm 4, in order to evaluate $\|(\mathrm{S}-2\,\mathrm{I})^{-1}\,\mathrm{S}\,\|_*$, given q^{n-1}, we need to calculate z^n, such that

$$z^n = (\mathrm{S}-2\,\mathrm{I})^{-1}\,\mathrm{S}\,q^{n-1}, \tag{26}$$

which is the solution to the following least square problem:

$$z^n = \arg\min_z \frac{1}{2}\|(\mathrm{S}-2\,\mathrm{I})z - \mathrm{S}\,q^{n-1}\|^2. \tag{27}$$

We apply gradient descent to solve (27):

$$\begin{aligned} z_{k+1}^n &= z_k^n - dt(\mathrm{S}-2\,\mathrm{I})[(\mathrm{S}^{\mathrm{T}}-2\,\mathrm{I})z_k^n - \mathrm{S}^{\mathrm{T}}\,q^{n-1}], \\ k &= 1, 2, 3, ..., K, \end{aligned} \tag{28}$$

where $z_1^n = z^{n-1}$, $z^n = z_{K+1}^n$. Besides, by substituting $z^N = (\mathrm{S}-2\,\mathrm{I})^{-1}\,\mathrm{S}\,q^{N-1}$, we have

$$\lambda^N = (q^N)^{\mathrm{T}}(\mathrm{S}-2\,\mathrm{I})^{-1}\,\mathrm{S}\,q^N = \langle q^N, z^{N+1}\rangle. \tag{29}$$

We summarize this modified power iterative method in Algorithm 5. Algorithm 5 extends the existing Algorithm 4 to evaluate the spectral norm of the multiplication of an inverse matrix $(\mathrm{S}-2\,\mathrm{I})^{-1}$ and S. By Algorithm 4, we are able to minimize the loss in (25).

For the parameters in Algorithms 4-5, we select $N = 10, K = 10, dt = 0.1, \epsilon = 0.1$, and $r = 10^{-3}$ to ensure the regularity conditions in (13) and (14).

Algorithm 5 Modified power iterative method

Given q^0 with $\|q^0\| = 1, \mathrm{S}, N, K, dt, z^0$
for $n = 1 : N$ **do**
  $z_1^n = z^{n-1}$
  for $k = 1 : K$ **do**
    $z_{k+1}^n = z_k^n - dt(\mathrm{S}-2\,\mathrm{I})[(\mathrm{S}^{\mathrm{T}}-2\,\mathrm{I})z_k^n - \mathrm{S}^{\mathrm{T}}\,q^{n-1}]$
  end for
  $z^n = z_{K+1}^n$
  $q^n = \frac{z^n}{\|z^n\|}$
end for
Return $\lambda^N = \langle q^N, z^{N+1}\rangle$

5. Experiments

In this section, we learn pseudo-contractive denoiser and k-strictly according to (25) and (21) with $k = \frac{1}{2}$. We use CBSD68 (Roth & Black, 2005) and Set12 (Zhang et al., 2017a) as test sets to show the effectiveness of our method. All the experiments are conducted under Linux system, Python 3.8.12 and Pytorch 1.10.2.

Training details. For the pseudo-contractive Gaussian denoisers, we select DRUNet (Zhang et al., 2021), which combines a residual learning (He et al., 2016) and UNet architecture (Ronneberger et al., 2015). DRUNet takes the noisy image, as well as the noise level σ as input, which is convenient for PnP image restoration.

For training details, we collect 800 images from DIV2K (Ignatov et al., 2019) as the training set and crop them into small patches of size 64×64. The batch size is 32. We add the Gaussian noise with $[\sigma_{\min}, \sigma_{\max}] = [0, 60]$ to the clean image. Adam optimizer is applied to train the model with learning rate $lr = 10^{-4}$. We set $r = 10^{-3}$ to ensure the regularity conditions in (13) and (14).

Denoising performance. We evaluate the Gaussian denoising performances of the proposed pseudo-contractive DRUNet (PC-DRUNet), $\frac{1}{2}$-strictly pseudo-contractive DRUNet (SPC-DRUNet), the non-expansive DRUNet (NE-DRUNet) trained with the loss (21) with $k = 0$, maximally monotone operator (MMO) (Pesquet et al., 2021) which is firmly non-expansive, Prox-DRUNet (Hurault et al., 2022b) with a contractive residual, the standard DRUNet without extra regularizations, the classical FFDNet (Zhang et al., 2018) and DnCNN (Zhang et al., 2017a). For a fair comparison, all denoisers are trained with DIV2K, and the patch sizes are set to 64. The PSNR values are given in Table 1 on CBSD68. The denoising performance on the gray images (Set12) is provided in Table 5 in Appendix J.

As shown in Table 1, restrictive conditions on the denoisers results in a compromised denoising performance. It can be explained by the spectrum distributions shown in Fig. 1: (a) for MMO; (b) for NE-DRUNet; (c) for Prox-DRUNet; (d) for SPC-DRUNet; (e) for PC-DRUNet; the complex plane $\mathbb{C}$ for DRUNet. A larger region means less restrictions on the Jacobian, and therefore, the denoising performance becomes

Table 1. Average denoising PSNR performance of different denoisers on CBSD68 dataset, for various noise levels σ.

σ	15	25	40
FFDNet	33.86	31.18	28.81
DnCNN	33.88	31.20	28.89
DRUNet	**34.14**	**31.54**	**29.33**
MMO	32.74	30.20	28.25
NE-DRUNet	32.97	30.54	28.50
Prox-DRUNet	33.18	30.60	28.38
SPC-DRUNet	34.12	31.51	29.32
PC-DRUNet	**34.14**	**31.53**	**29.32**

better. In Fig. 1, we have $(a) \subset (b) \subset (d) \subset (e) \subset \mathbb{C}$, and the PSNR values in Table 1 by MMO, NE-DRUNet, SPC-DRUNet, PC-DRUNet, and DRUNet have the same order. It indicates that pseudo-contractiveness is a weaker and less harmful assumption on the deep denoisers.

Assumption validations. In the experiments, the strictly pseudo-contractive and pseudo-contractive conditions are softly constrained by the loss functions (21) and (25) with a trade-off parameter r. We validate the conditions in Table 2. As shown in Table 2, DRUNet without spectral regularization term is neither non-expansive nor pseudo-contractive. When PC-DRUNet and SPC-DRUNet are trained by the loss functions (21) and (25) with $r = 10^{-3}$, the norms $\left\| \frac{1}{2}\mathrm{I} + \frac{1}{2}\mathrm{J} \right\|_*$ and $\|(\mathrm{S} - 2\,\mathrm{I})^{-1}\,\mathrm{S}\,\|_*$ are smaller than 1. It validates the effectiveness of the proposed training strategy.

Table 2. Maximal values of different norms on CBSD68 dataset for various noise levels σ.

σ	15	25	40	Max. Norm
DRUNet	9.544	10.81	13.16	$\|\,\mathrm{J}\,\|_*$
DRUNet	1.614	2.998	1.775	$\|\frac{1}{2}\mathrm{I} + \frac{1}{2}\mathrm{J}\,\|_*$
DRUNet	4.364	4.429	4.346	$\|(\mathrm{S} - 2\,\mathrm{I})^{-1}\,\mathrm{S}\,\|_*$
SPC-DRUNet ($r = 10^{-3}$)	0.995	0.991	0.999	$\|\frac{1}{2}\mathrm{I} + \frac{1}{2}\mathrm{J}\,\|_*$
SPC-DRUNet ($r = 10^{-4}$)	1.014	1.186	1.440	$\|\frac{1}{2}\mathrm{I} + \frac{1}{2}\mathrm{J}\,\|_*$
PC-DRUNet ($r = 10^{-3}$)	0.988	0.999	0.996	$\|(\mathrm{S} - 2\,\mathrm{I})^{-1}\,\mathrm{S}\,\|_*$
PC-DRUNet ($r = 10^{-4}$)	1.020	1.001	1.246	$\|(\mathrm{S} - 2\,\mathrm{I})^{-1}\,\mathrm{S}\,\|_*$

PnP restoration. We apply the proposed PnPI-GD, PnPI-HQS, and PnPI-FBS algorithms on deblurring and super-resolution tasks. In Appendix O, we consider poisson denoising experiments. In Appendix P, we also apply PnPI-HQS to traffic data completion task. For PnPI-GD, we choose the pretrained PC-DRUNet as D_σ. For PnPI-HQS and PnPI-FBS, we choose the pretrained SPC-DRUNet with $k = \frac{1}{2}$ according to remark 3.7. We apply a decreasing step size strategy in PnPI-HQS, by multiplying β by a factor ρ slightly bigger than 1, and multiplying σ by $\frac{1}{\sqrt{\rho}}$ in each iteration as suggested in (Zhang et al., 2021).

For the step size sequences $\{\alpha_n\}, \{\beta_n\}$ in Algorithms 1-3, we let $\alpha_n = (n+1)^{-a}, \beta_n = (n+1)^{-b}$, with $0 < b < a < 1, a + b < 1$, to satisfy the condition (20) in Theorem 3.3. Large a, b leads to small step size. Emprically, we let $a = 0.3, b = 0.15$ in PnPI-GD and PnPI-FBS, and $a = 0.8, b = 0.15$ in PnPI-HQS. In the deblurring task, the proposed methods are initialized with the observed image, that is $u^0 = f$. In the single image super-resolution task, we choose u^0 as the bicubic interpolation of f as in (Zhang et al., 2021).

We compare our methods with some state-of-the-art convergent PnP methods including MMO-FBS (Pesquet et al., 2021), which uses the FBS method with MMO denoiser; NE-PGD (Reehorst & Schniter, 2018; Liu et al., 2021) using PGD framework with NE-DRUNet; Prox-DRS (Hurault et al., 2022b), which uses the Douglas-Rachfold Spilitting (DRS) method with Prox-DRUNet. We also indicate the results by DPIR (Zhang et al., 2021), which applies PnP-HQS method with decreasing step size and DRUNet denoiser, but without convergence guarantee.

Deblurring. In the deblurring task, we seek to solve the inverse problem (1) with a convolution operator K performed with circular boundary conditions and Gaussian noise n with zero mean value and standard derivation σ. The fidelity term is $G(u; f) = \frac{\mu}{2}\|Ku - f\|^2$, where $\mu > 0$ is the balancing parameter. The proximal operator $\mathrm{Prox}_{\frac{G}{\beta}}$ can be efficiently calculated as in (Pan et al., 2016). Note that in this case, $\nabla G(u) = \mu K^{\mathrm{T}}(Ku - f)$, and $\|K^{\mathrm{T}}K\|_* \leq 1$. Therefore, ∇G is μ-cocoercive. We set $N = 100$ for PnPI-GD, $N = 50$ for PnPI-FBS, $N = 25$ for PnPI-HQS, and fine tune μ, λ and $\beta > 0$ in the proposed methods to achieve the best quantitive PSNR values.

We demonstrate the effectiveness of our methods on the 8 real-world camera shake kernels by Levin et al. (Levin et al., 2009), with $\sigma = 12.75$, and 17.85 respectively. The kernels are shown in Appendix K, and deblurring results on Set12 are provided in Appendix M.

We summarize the PSNR and SSIM values with $\sigma = 12.75$ and 17.85 in Table 3. The detailed PSNR and SSIM values are listed in Appendix L. The highest value is marked in **boldface**. It can be seen that on average, PnPI-HQS provides the best PSNR and SSIM values. Compared with the convergent PnP methods MMO-FBS, NE-PGD, and Prox-DRS, the proposed PnPI-GD and PnPI-FBS provide competitive results. It validates the effectiveness of PnP Ishikawa scheme and the pseudo-contractive denoisers.

In Fig. 2, we show deblurring results when recovering the image '0037' from CBSD68 with kernel 2 and Gaussian noise $\sigma = 12.75$. It can be seen in Fig. 2 (a) that the image is severely blurred and noisy. Compared with MMO-

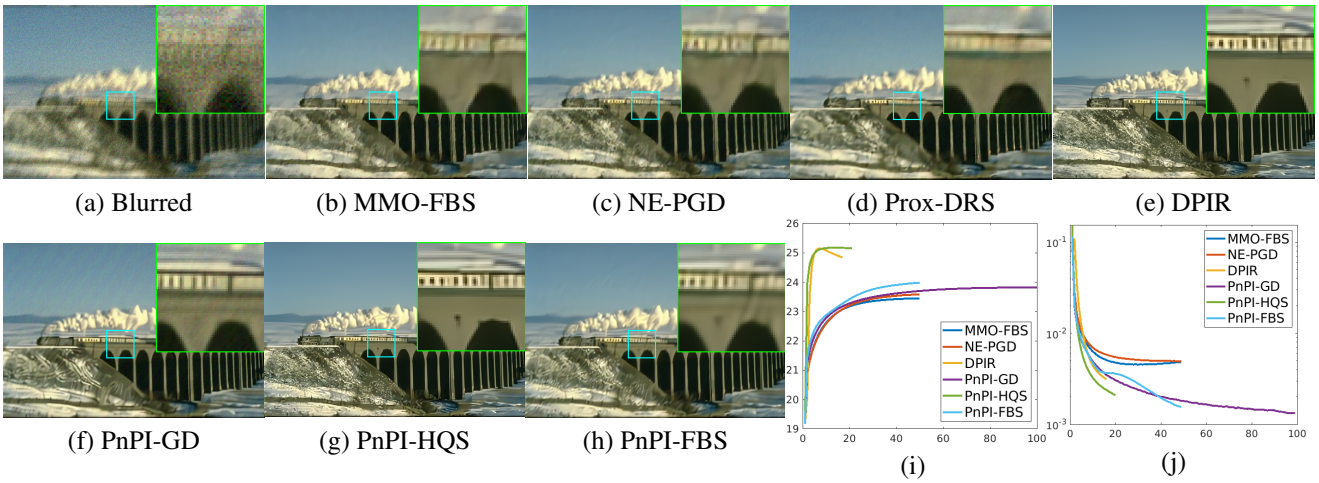


Figure 2. Results by different methods when recovering the image '0037' from CBSD68 with kernel 2 and Gaussian noise with $\sigma = 12.75$. (a) Blurred. (b) MMO-FBS, PSNR=23.45dB. (c) NE-PGD, PSNR=23.59dB. (d) Prox-DRS, PSNR=23.53dB. (e) DPIR, PSNR=24.85dB. (f) PnPI-GD, PSNR=23.82dB. (g) PnPI-HQS, PSNR=25.16dB. (h) PnPI-FBS, PSNR=23.99dB. (i) PSNR curves, x-axis denotes iteration number. (j) Relative error curves, x-axis denotes iteration number.

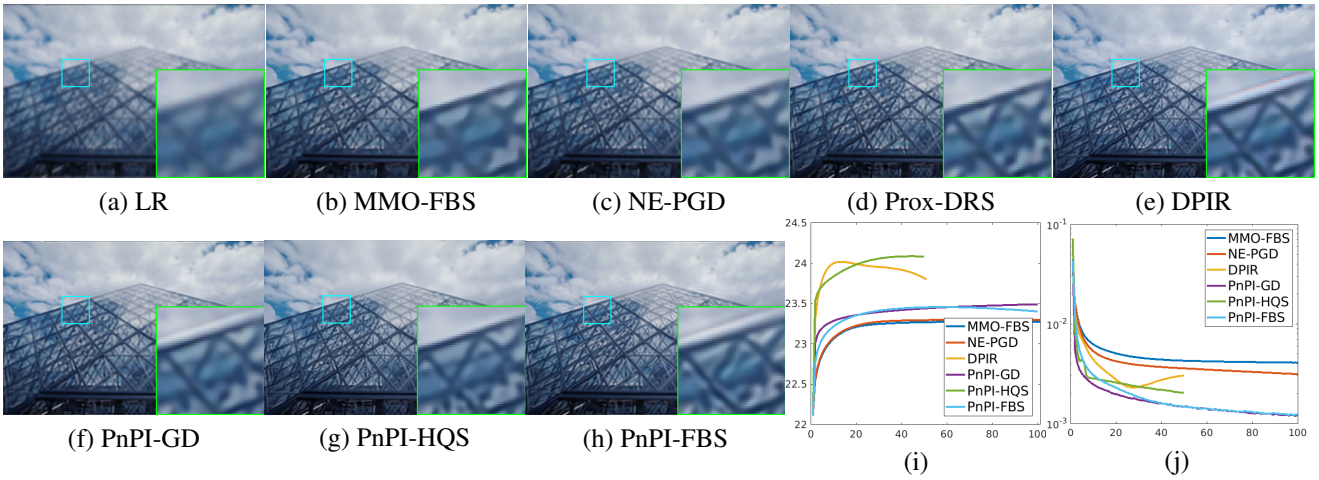


Figure 3. Super-resolution results by different methods on the image '0046' from CBSD68 with $s = 2, \sigma = 2.55$. (a) Low-resolution (LR). (b) MMO-FBS, PSNR=23.28dB. (c) NE-PGD, PSNR=23.30dB. (d) Prox-DRS, PSNR=23.62dB. (e) DPIR, PSNR=23.80dB. (f) PnPI-GD, PSNR=23.50dB. (g) PnPI-HQS, PSNR=24.08dB. (h) PnPI-FBS, PSNR=23.38dB. (i) PSNR curves, x-axis denotes iteration number. (j) Relative error curves, x-axis denotes iteration number.

FBS and NE-PGD, PnPI-GD and PnPI-FBS provide results with better structure recovery, see Figs. 2 (b) (c) (f) (h). Compared with Prox-DRS and DPIR, the result by PnPI-HQS has clearer details, see Figs. 2 (d) (e) (g). Note that DPIR has no convergence guarantee, while PnPI-GD, PnPI-HQS, and PnPI-FBS are shown convergent in Figs. 2 (i) (j).

Table 3. Average deblurring PSNR and SSIM performance by different methods on CBSD68 dataset with Levin's 8 kernels with $\sigma = 12.75$ and 17.85.

	$\sigma = 12.75$		$\sigma = 12.75$	
	PSNR	SSIM	PSNR	SSIM
MMO-FBS	26.35	0.7100	25.72	0.7000
NE-PGD	26.58	0.7277	25.94	0.6983
Prox-DRS	26.64	0.7200	25.99	0.6900
DPIR	27.65	0.7738	26.75	0.7293
PnPI-GD	26.41	0.6962	25.61	0.6633
PnPI-HQS	**27.75**	**0.7797**	**26.77**	**0.7386**
PnPI-FBS	26.83	0.7451	25.97	0.7081

Single image super-resolution. In the super-resolution task, we seek to solve the inverse problem (1) with a convolution operator K performed with circular boundary conditions, a standard s-fold downsampling operator S, as well as Gaussian noise n with zero mean value and standard derivation σ. The fidelity term is $G(u; f) = \frac{\mu}{2}\|SKu - f\|^2$, where $\mu > 0$ is the balancing parameter. The proximal operator $\mathrm{Prox}_{\frac{G}{\beta}}$ can be efficiently calculated as in (Zhang et al., 2021). Similarly to the deblurring task, $\nabla G(u) = \mu S^{\mathrm{T}} K^{\mathrm{T}}(Ku - f)$, and $\|S^{\mathrm{T}} K^{\mathrm{T}} KS\|_* \leq 1$. Therefore, ∇G is μ-cocoercive. We set $N = 100$ for PnPI-GD and PnPI-FBS, $N = 50$ for PnPI-HQS, and fine tune μ, λ and $\beta > 0$ to achieve the best quantitive PSNR values.

We let the kernel K be the isotropic Gaussian blur kernel with standard deviation 2. The downsampling scale are set as $s = 2, 4$. The noise levels are set as $\sigma = 0, 2.55, 7.65$.

We summarize the PSNR and SSIM values in Table 4. It can be seen that in most cases, PnPI-HQS provides the best PSNR and SSIM values. Compared with the convergent PnP methods, PnPI-GD and PnPI-FBS provides competitive results, especially when the degradation is severe. Results on Set12 are provided in Appendix N.

In Fig. 3, we provide visual results on the image '0046' when $s = 2$ and $\sigma = 2.55$. In Figs. 3 (b) (c) (f) (h), compared with MMO-FBS and NE-PGD, the proposed PnPI-GD and PnPI-FBS provide sharper edges. PnPI-HQS has clearer structures than Prox-DRS, see Figs. 3 (d) and (g). Note that the result by DPIR seems have some ringing artifacts. We account this for the non-convergent behavior of DPIR shown in Figs. 3 (i) (j), while the proposed methods have stable and convergent PSNR and relative error curves.

Table 4. Average super-resolution PSNR and SSIM performance by different methods on CBSD68 dataset with different scales and noise levels.

scale	s=2			s=4		
σ	0	2.55	7.65	0	2.55	7.65
MMO-FBS	27.02	26.16	25.28	25.30	25.17	24.51
	0.7719	0.7142	0.6604	0.6692	0.6602	0.6285
NE-PGD	27.02	26.23	25.27	25.34	25.21	24.54
	0.7822	0.7197	0.6622	0.6719	0.6632	0.6311
Prox-DRS	30.26	26.63	25.57	25.49	25.23	24.48
	0.8874	0.7364	0.6805	0.7007	0.6716	0.6280
DPIR	29.95	27.06	25.77	25.82	25.42	24.68
	0.8677	**0.7615**	0.6993	0.7099	0.6808	0.6398
PnPI-GD	25.54	25.36	25.12	25.06	25.01	24.30
	0.7201	0.7028	0.6766	0.6838	0.6728	0.6280
PnPI-HQS	**30.38**	**27.09**	**25.93**	**25.83**	**25.47**	**24.76**
	0.8822	0.7602	**0.7012**	**0.7108**	**0.6878**	**0.6480**
PnPI-FBS	28.12	26.29	25.44	25.43	25.29	24.53
	0.8306	0.7357	0.6846	0.6877	0.6795	0.6380

6. Conclusion

This paper introduces a novel training strategy that enforces a weaker constraint on the deep denoiser called pseudo-contractiveness. By studying the spectrum of the Jacobian matrix, we uncover relationships between different denoiser assumptions. Utilizing the Ishikawa process, efficient fixed-point algorithms are derived. The proposed algorithms demonstrate strong theoretical convergence towards a fixed point. To enforce the pseudo-contractive denoiser assumption, a training strategy based on holomorphic transformation and functional calculi is proposed. Extensive experiments showcase the superior performance of the pseudo-contractive denoiser compared to other related denoisers, both visually and quantitatively. Overall, the proposed methods offer competitive results for image restoration tasks.

Source code

The source code and pretrained models are available at https://github.com/FizzzFizzz/Learning-Pseudo-Contractive-Denoisers-for-Inverse-Problems.

Acknowledgments

We sincerely thank the anonymous reviewers for the suggestions. We thank Professor Hang Wang for her helpful and inspiring discussions on the spectral mapping theorem and functional calculi. This work is supported by National Key R&D Program of China (Nos. 2021YFA1000300 and 2021YFA1000302), Fundamental Research Funds for the Central Universities, Natural Science Foundation of Shanghai (No. 22ZR1419500), Science and Technology Commission of Shanghai Municipality (No. 22DZ2229014). and Natural Science Foundation of Chongqing, China (No. CSTB2023NSCQ-MSX0276),

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Bauschke, H. H., Combettes, P. L., Bauschke, H. H., and Combettes, P. L. *Correction to: Convex analysis and monotone operator theory in hilbert spaces*. Springer, 2017.

Beck, A. *First-order methods in optimization*. SIAM, 2017.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Bredies, K., Kunisch, K., and Pock, T. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

Buades, A., Coll, B., and Morel, J.-M. A review of image denoising algorithms, with a new one. *Multiscale modeling & simulation*, 4(2):490–530, 2005.

Buzzard, G. T., Chan, S. H., Sreehari, S., and Bouman, C. A. Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium. *SIAM Journal on Imaging Sciences*, 11(3), 2018.

Chan, S. H., Wang, X., and Elgendy, O. A. Plug-and-play admm for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98, 2016.

Chen, P., Li, F., Wei, D., and Lu, C. Spatiotemporal traffic data completion with truncated minimax-concave penalty. *Transportation Research Part C: Emerging Technologies*, 164:104657, 2024.

Chen, X., He, Z., Chen, Y., Lu, Y., and Wang, J. Missing traffic data imputation and pattern discovery with a bayesian augmented tensor factorization model. *Transportation Research Part C: Emerging Technologies*, 104: 66–77, 2019a.

Chen, X., He, Z., and Sun, L. A bayesian tensor decomposition approach for spatiotemporal traffic data imputation. *Transportation research part C: emerging technologies*, 98:73–84, 2019b.

Chen, X., Yang, J., and Sun, L. A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation. *Transportation Research Part C: Emerging Technologies*, 117:102673, 2020.

Chen, X., Lei, M., Saunier, N., and Sun, L. Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):12301–12310, 2021.

Chidume, C. Iterative approximation of fixed points of lipschitzian strictly pseudocontractive mappings. *Proceedings of the American Mathematical Society*, 99(2): 283–288, 1987.

Cohen, R., Blau, Y., Freedman, D., and Rivlin, E. It has potential: Gradient-driven denoisers for convergent solutions to inverse problems. *Advances in Neural Information Processing Systems*, 34:18152–18164, 2021.

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. Image denoising with block-matching and 3d filtering. In *Image processing: algorithms and systems, neural networks, and machine learning*, volume 6064, pp. 354–365. SPIE, 2006.

Giselsson, P. Tight global linear convergence rate bounds for douglas–rachford splitting. *Journal of Fixed Point Theory and Applications*, 19(4):2241–2270, 2017.

Golub, G. H. and Van Loan, C. F. *Matrix computations*. JHU press, 2013.

Gu, S., Zhang, L., Zuo, W., and Feng, X. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2862–2869, 2014.

Haase, M. Spectral mapping theorems for holomorphic functional calculi. *Journal of the London Mathematical Society*, 71(3):723–739, 2005.

Harte, R. Spectral mapping theorems. In *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences*, pp. 89–107. JSTOR, 1972.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Heide, F., Steinberger, M., Tsai, Y.-T., Rouf, M., Pajak, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., et al. Flexisp: A flexible camera image processing framework. *ACM Transactions on Graphics (ToG)*, 33 (6):1–13, 2014.

Hicks, T. L. and Kubicek, J. D. On the mann iteration process in a hilbert space. 1977.

Hurault, S., Leclaire, A., and Papadakis, N. Gradient step denoiser for convergent plug-and-play. In *International Conference on Learning Representations (ICLR'22)*, 2022a.

Hurault, S., Leclaire, A., and Papadakis, N. Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. In *International Conference on Machine Learning*, pp. 9483–9505. PMLR, 2022b.

Ignatov, A., Timofte, R., et al. Pirm challenge on perceptual image enhancement on smartphones: report. In *European Conference on Computer Vision (ECCV) Workshops*, January 2019.

Ishikawa, S. Fixed points by a new iteration method. *Proceedings of the American Mathematical Society*, 44(1): 147–150, 1974.

Kumar, P. G. and Ranjan Sahay, R. Low rank poisson denoising (lrpd): A low rank approach using split bregman algorithm for poisson noise removal from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0, 2019.

Laroche, C., Almansa, A., Coupeté, E., and Tassano, M. Provably convergent plug & play linearized admm, applied to deblurring spatially varying kernels. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023a.

Laroche, C., Almansa, A., and Tassano, M. Deep model-based super-resolution with non-uniform blur. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1797–1808, 2023b.

Laumont, R., De Bortoli, V., Almansa, A., Delon, J., Durmus, A., and Pereyra, M. On maximum a posteriori estimation with plug & play priors and stochastic gradient descent. *Journal of Mathematical Imaging and Vision*, 65(1):140–163, 2023.

Le Pendu, M. and Guillemot, C. Preconditioned plug-and-play admm with locally adjustable denoiser for image restoration. *SIAM Journal on Imaging Sciences*, 16(1): 393–422, 2023.

Levin, A., Weiss, Y., Durand, F., and Freeman, W. T. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 1964–1971. IEEE, 2009.

Liu, J., Asif, S., Wohlberg, B., and Kamilov, U. Recovery analysis for plug-and-play priors using the restricted eigenvalue condition. *Advances in Neural Information Processing Systems*, 34:5921–5933, 2021.

Mairal J, Bach F, P. J. e. a. Non-local sparse models for image restoration. *IEEE 12th International Conference on Computer Vision*, pp. 2272–2279, 2009.

Nie, T., Qin, G., and Sun, J. Truncated tensor schatten p-norm based approach for spatiotemporal traffic data imputation with complicated missing patterns. *Transportation research part C: emerging technologies*, 141: 103737, 2022.

Pan, J., Hu, Z., Su, Z., and Yang, M.-H. l_0-regularized intensity and gradient prior for deblurring text images and beyond. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):342–355, 2016.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in pytorch. 2017.

Peng, M., Kitichotkul, R., Seidel, S. W., Yu, C., and Goyal, V. K. Denoising particle beam micrographs with plug-and-play methods. *IEEE Transactions on Computational Imaging*, 2023.

Pesquet, J.-C., Repetti, A., Terris, M., and Wiaux, Y. Learning maximally monotone operators for image recovery. *SIAM Journal on Imaging Sciences*, 14(3):1206–1237, 2021.

Rafiq, A. On mann iteration in hilbert spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 66(10):2230–2236, 2007.

Reehorst, E. T. and Schniter, P. Regularization by denoising: Clarifications and new interpretations. *IEEE transactions on computational imaging*, 5(1):52–67, 2018.

Romano, Y., Elad, M., and Milanfar, P. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241. Springer, 2015.

Roth, S. and Black, M. J. Fields of experts: A framework for learning image priors. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pp. 860–867. IEEE, 2005.

Rudin, L. I., Osher, S., and Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

Ryu, E., Liu, J., Wang, S., Chen, X., Wang, Z., and Yin, W. Plug-and-play methods provably converge with properly trained denoisers. In *International Conference on Machine Learning*, pp. 5546–5557. PMLR, 2019.

Sreehari, S., Venkatakrishnan, S. V., Wohlberg, B., Buzzard, G. T., Drummy, L. F., Simmons, J. P., and Bouman, C. A. Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Transactions on Computational Imaging*, 2(4):408–423, 2016.

Sun, Y., Wohlberg, B., and Kamilov, U. S. An online plug-and-play algorithm for regularized image reconstruction. *IEEE Transactions on Computational Imaging*, 5(3):395–408, 2019.

Tirer, T. and Giryes, R. Image restoration by iterative denoising and backward projections. *IEEE Transactions on Image Processing*, 28(3):1220–1234, 2018.

Venkatakrishnan, S. V., Bouman, C. A., and Wohlberg, B. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pp. 945–948. IEEE, 2013.

Wei, D., Li, F., and Weng, S. Cauchy noise removal via convergent plug-and-play framework with outliers detection. *Journal of Scientific Computing*, 96(3):76, 2023a.

Wei, D., Weng, S., and Li, F. Nonconvex rician noise removal via convergent plug-and-play framework. *Applied Mathematical Modelling*, 123:197–212, 2023b.

Weng, X. Fixed point iteration for local strictly pseudo-contractive mapping. *Proceedings of the American Mathematical Society*, 113(3):727–731, 1991.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017a.

Zhang, K., Zuo, W., Gu, S., and Zhang, L. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3929–3938, 2017b.

Zhang, K., Zuo, W., and Zhang, L. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L., and Timofte, R. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021.

Zhu, Y., Zhang, K., Liang, J., Cao, J., Wen, B., Timofte, R., and Van Gool, L. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1219–1229, 2023.

A. Closely related assumptions

We briefly review some closely related assumptions. Let V be the real Hilbert space, $\langle \cdot, \cdot \rangle$ be the inner product on V, and $\| \cdot \|$ be the induced norm.

• Non-expansive D:

$$\| \mathrm{D}(x) - \mathrm{D}(y) \| \leq \|x - y\|, \forall x, y \in V. \tag{30}$$

• θ-averaged D ($\theta \in (0, 1]$):

$$\left\| \left[\left(1 - \frac{1}{\theta}\right) \mathrm{I} + \frac{1}{\theta} \mathrm{D} \right](x) - \left[\left(1 - \frac{1}{\theta}\right) \mathrm{I} + \frac{1}{\theta} \mathrm{D} \right](y) \right\| \leq \|x - y\|, \forall x, y \in V. \tag{31}$$

• Contractive $\mathrm{I} - \mathrm{D}$ ($r < 1$):

$$\|(\mathrm{I} - \mathrm{D})(x) - (\mathrm{I} - \mathrm{D})(y)\| \leq r\|x - y\|, \forall x, y \in V. \tag{32}$$

θ-averaged D can be written as

$$\mathrm{D} = \theta \, \mathrm{N} + (1 - \theta) \, \mathrm{I}, \tag{33}$$

where N is a non-expansive mapping. Averaged mappings are non-expansive. Firmly non-expansiveness is a special case of averagedness with $\theta = \frac{1}{2}$.

We give the *regularity conditions* for the Jacobian J under different assumptions on the denoiser D, as well as the distribution regions $\mathrm{Sp}(\mathrm{J})$ in (34)-(38). $\| \cdot \|_*$ denotes the spectral norm. Note that these regularity conditions are sufficient conditions for a denoiser to satisfy the corresponding assumptions.

• Non-expansive D:

$$\| \mathrm{J} \|_* \leq 1, \ \mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : |z| \leq 1\}. \tag{34}$$

• θ-averaged D ($\theta \in (0, 1]$):

$$\left\| \left[\left(1 - \frac{1}{\theta}\right) \mathrm{I} + \frac{1}{\theta} \mathrm{J} \right] \right\|_* \leq 1, \ \mathrm{Sp}(\mathrm{J}) \subset \left\{ z \in \mathbb{C} : \left| 1 - \frac{1}{\theta} + \frac{1}{\theta} z \right| \leq 1 \right\}. \tag{35}$$

• Contractive $\mathrm{I} - \mathrm{D}$ ($r < 1$):

$$\| \mathrm{I} - \mathrm{J} \|_* \leq 1, \mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : |z - 1| \leq r\}. \tag{36}$$

• k-strictly pseudo-contractive D ($k < 1$):

$$\|k \, \mathrm{I} + (1 - k) \, \mathrm{J} \|_* \leq 1, \ \mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : |(1 - k)z + k| \leq 1\}. \tag{37}$$

• Pseudo-contractive D:

$$\langle (\mathrm{I} - \mathrm{J}^{\mathrm{T}})(x - y), x - y \rangle \geq 0, \ \mathrm{Sp}(\mathrm{J}) \subset \{z \in \mathbb{C} : real(z) \leq 1\}. \tag{38}$$

In Fig. 1, we present an intuitive illustration of the relationships between different assumptions on D. We display the regions of spectrum distribution on the complex plane. In (a), we depict the region when D is $\frac{1}{2}$-averaged, which corresponds to firm non-expansiveness. (a) is contained within panel (b), representing the unit disk when D is non-expansive. This reveals that firm non-expansiveness is a specific instance of non-expansiveness. (c) showcases the region where $\mathrm{I} - \mathrm{D}$ is contractive with $r = \frac{1}{2}$. In Figs. 1 (d) and (e), we plot the distribution region for $\frac{1}{2}$-strictly pseudo-contractive and pseudo-contractive D, respectively. The are in (d) encompasses non-expansiveness and is enclosed by the half-plane in (e). This suggests that the (strictly) pseudo-contractive property constitutes a significantly weaker assumption for denoisers.

B. Proofs of Lemmas and Theorems

Before the proofs, we review Lemma B.1 from (Giselsson, 2017).

Lemma B.1. *Let G be proper, closed, and convex, ∇G is γ-cocoercive, that is for any $x, y \in V$, there holds*

$$\langle x - y, \nabla G(x) - \nabla(y) \rangle \geq \gamma \|\nabla G(x) - \nabla G(y)\|^2. \tag{39}$$

Then, the resolvent of ∇G, which is the proximal operator $\mathrm{P} = \mathrm{Prox}_G = (\mathrm{I} + \nabla G)^{-1}$, is $\frac{1}{2\gamma+2}$-averaged. The reflective resolvent of ∇G, $2\,\mathrm{P} - \mathrm{I} = 2(\mathrm{I} + \nabla G)^{-1} - \mathrm{I}$ is $\frac{1}{1+\gamma}$-averaged.

C. Proof of Lemma 2.1

Proof. By the definition, D is said to be k-strictly pseudo-contractive with $k < 1$, if $\forall x, y \in V$, we have

$$\| \mathrm{D}(x) - \mathrm{D}(y)\|^2 \leq \|x - y\|^2 + k\|(\mathrm{I} - \mathrm{D})(x) - (\mathrm{I} - \mathrm{D})(y)\|^2. \tag{40}$$

Denote $a = \mathrm{D}(x) - \mathrm{D}(y), b = x - y$. Then

$$\begin{aligned}
&\|a\|^2 && \leq \|b\|^2 + k\|a - b\|^2 = \|b\|^2 + k\|a\|^2 + k\|b\|^2 - 2k\langle a, b\rangle, \\
&(1-k)\|a\|^2 + 2k\langle a, b\rangle && \leq (1+k)\|b\|^2, \\
&\|a\|^2 + \frac{2k}{1-k}\langle a, b\rangle && \leq \frac{1+k}{1-k}\|b\|^2, \\
&\left\|a + \frac{k}{1-k}b\right\|^2 && \leq \left(\frac{k^2}{(1-k)^2} + \frac{1+k}{1-k}\right)\|b\|^2 = \left(\frac{1}{(1-k)^2}\right)\|b\|^2, \\
&\|(1-k)a + kb\|^2 && \leq \|b\|,
\end{aligned} \tag{41}$$

which means that $(1-k)\,\mathrm{D} + k\,\mathrm{I}$ is non-expansive. Let $\mathrm{N} = (1-k)\,\mathrm{D} + k\,\mathrm{I}$, we have

$$\mathrm{D} = \frac{1}{1-k}\,\mathrm{N} - \frac{k}{1-k}\,\mathrm{I}\,. \tag{42}$$

□

D. Proof of Lemma 2.2

Proof.

$$\begin{aligned}
& \mathrm{D} \text{ is pseudo-contractive} \\
\iff\ & \forall x, y \in V,\ \| \mathrm{D}(x) - \mathrm{D}(y)\|^2 \leq \|x - y\|^2 + \|(\mathrm{I} - \mathrm{D})(x) - (\mathrm{I} - \mathrm{D})(y)\|^2 \\
\iff\ & \forall x, y \in V,\ \|(\mathrm{I} - \mathrm{D} - \mathrm{I})(x) - (\mathrm{I} - \mathrm{D} - \mathrm{I})(y)\|^2 \leq \|x - y\|^2 + \|(\mathrm{I} - \mathrm{D})(x) - (\mathrm{I} - \mathrm{D})(y)\|^2 \\
\iff\ & \forall x, y \in V,\ \|(\mathrm{I} - \mathrm{D})(x) - (\mathrm{I} - \mathrm{D})(y)\|^2 + \|x - y\|^2 - 2\langle(\mathrm{I} - \mathrm{D})(x) - (\mathrm{I} - \mathrm{D})(y), x - y\rangle \leq \|x - y\|^2 \\
& \qquad\qquad + \|(\mathrm{I} - \mathrm{D})(x) - (\mathrm{I} - \mathrm{D})(y)\|^2 \\
\iff\ & \forall x, y \in V,\ \langle(\mathrm{I} - \mathrm{D})(x) - (\mathrm{I} - \mathrm{D})(y), x - y\rangle \geq 0.
\end{aligned} \tag{43}$$

□

E. Proof of Lemma 3.1

Proof. According to Lemma 2.2, we only need to show that for any $x, y \in V$, there holds

$$\langle(\mathrm{I} - \mathrm{T})(x) - (\mathrm{I} - \mathrm{T})(y), x - y\rangle \geq 0. \tag{44}$$

Note that $\mathrm{I} - \mathrm{T} = \mathrm{I} - \mathrm{D} + \nabla G$. Thus we have

$$\langle(\mathrm{I} - \mathrm{T})(x) - (\mathrm{I} - \mathrm{T})(y), x - y\rangle = \langle(\mathrm{I} - \mathrm{D})(x) - (\mathrm{I} - \mathrm{D})(y), x - y\rangle + \langle\nabla G(x) - \nabla G(y), x - y\rangle \geq 0 + 0 = 0. \tag{45}$$

The last $\geq$ comes from the pseudo-contractive D and convex G. □

F. Proof of Lemma 3.2

Before the proof of Lemma 3.2, we give the following Lemma F.1.

Lemma F.1. *Let V be the real Hilbert space. For any $x, y \in V$ and $\alpha, \beta \in \mathbb{R}$, there holds*

$$\|\alpha x + \beta y\|^2 = \alpha(\alpha + \beta)\|x\|^2 + \beta(\alpha + \beta)\|y\|^2 - \alpha\beta\|x - y\|^2, \tag{46}$$

and

$$\alpha\beta\|x + y\|^2 = \alpha(\alpha + \beta)\|x\|^2 + \beta(\alpha + \beta)\|y\|^2 - \|\alpha x - \beta y\|^2. \tag{47}$$

Proof. Here we only prove the first equality. By letting $x' = x, y' = -y$, the second equality holds naturally. The left hand side equals to

$$LHS = \alpha^2\|x\|^2 + \beta^2\|y\|^2 + 2\alpha\beta\langle x, y\rangle, \tag{48}$$

while the right hand side equals to

$$\begin{aligned} RHS &= \alpha^2\|x\|^2 + \beta^2\|y\|^2 + \alpha\beta(\|x\|^2 + \|y\|^2) - \alpha\beta(\|x\|^2 + \|y\|^2 - 2\langle x, y\rangle) \\ &= \alpha^2\|x\|^2 + \beta^2\|y\|^2 + 2\alpha\beta\langle x, y\rangle = LHS. \end{aligned} \tag{49}$$

□

Now we can prove Lemma 3.2.

Proof. Since D is k-strictly pseudo-contractive, and that P is θ-averaged, for any $x, y \in V$ we have

$$\begin{aligned} &\|\mathrm{D}\circ\mathrm{P}(x) - \mathrm{D}\circ\mathrm{P}(y)\|^2 \le \|\mathrm{P}(x) - \mathrm{P}(y)\|^2 + k\|(\mathrm{I}-\mathrm{D})\circ\mathrm{P}(x) - (\mathrm{I}-\mathrm{D})\circ\mathrm{P}(y)\|^2 \\ \le\ &\|x-y\|^2 - \tfrac{1-\theta}{\theta}\|(\mathrm{I}-\mathrm{P})(x) - (\mathrm{I}-\mathrm{P})(y)\|^2 + k\|(\mathrm{I}-\mathrm{D})\circ\mathrm{P}(x) - (\mathrm{I}-\mathrm{D})\circ\mathrm{P}(y)\|^2, \end{aligned} \tag{50}$$

where the second inequality comes from Proposition 4.35 in (Bauschke et al., 2017). Set

$$\alpha = -\frac{1-\theta}{\theta}, \beta = k, l = \frac{\alpha\beta}{\alpha+\beta} = \frac{k\frac{1-\theta}{\theta}}{\frac{1-\theta}{\theta} - k} = \frac{k(1-\theta)}{(1-\theta) - k\theta}. \tag{51}$$

By Lemma F.1, there holds

$$\begin{aligned} &\alpha\|(\mathrm{I}-\mathrm{P})(x) - (\mathrm{I}-\mathrm{P})(y)\|^2 + \beta\|(\mathrm{I}-\mathrm{D})\circ\mathrm{P}(x) - (\mathrm{I}-\mathrm{D})\circ\mathrm{P}(y)\|^2 \\ =\ &\frac{\alpha\beta}{\alpha+\beta}\|[(\mathrm{I}-\mathrm{P}) + (\mathrm{I}-\mathrm{D})\circ\mathrm{P}](x) - [(\mathrm{I}-\mathrm{P}) + (\mathrm{I}-\mathrm{D})\circ\mathrm{P}](y)\|^2 \\ &+\frac{1}{\alpha+\beta}\|\alpha(\mathrm{I}-\mathrm{D})\circ\mathrm{P}(x) - \alpha(\mathrm{I}-\mathrm{D})\circ\mathrm{P}(y) - \beta(\mathrm{I}-\mathrm{P})(x) + \beta(\mathrm{I}-\mathrm{P})(y)\|^2 \\ =\ &\frac{\alpha\beta}{\alpha+\beta}\|(\mathrm{I}-\mathrm{D}\circ\mathrm{P})(x) - (\mathrm{I}-\mathrm{D}\circ\mathrm{P})(y)\|^2 \\ &+\frac{1}{\alpha+\beta}\|\alpha(\mathrm{I}-\mathrm{D})\circ\mathrm{P}(x) - \alpha(\mathrm{I}-\mathrm{D})\circ\mathrm{P}(y) - \beta(\mathrm{I}-\mathrm{P})(x) + \beta(\mathrm{I}-\mathrm{P})(y)\|^2. \end{aligned} \tag{52}$$

When $k \le 1-\theta$, $\alpha+\beta < 0$, and thus

$$\begin{aligned} &\alpha\|(\mathrm{I}-\mathrm{P})(x) - (\mathrm{I}-\mathrm{P})(y)\|^2 + \beta\|(\mathrm{I}-\mathrm{D})\circ\mathrm{P}(x) - (\mathrm{I}-\mathrm{D})\circ\mathrm{P}(y)\|^2 \\ \le\ &\frac{\alpha\beta}{\alpha+\beta}\|(\mathrm{I}-\mathrm{D}\circ\mathrm{P})(x) - (\mathrm{I}-\mathrm{D}\circ\mathrm{P})(y)\|^2 \\ =\ &l\|(\mathrm{I}-\mathrm{D}\circ\mathrm{P})(x) - (\mathrm{I}-\mathrm{D}\circ\mathrm{P})(y)\|^2. \end{aligned} \tag{53}$$

If $k = 1-\theta$, $l = 1$. This completes the proof. □

G. Proof of Theorem 3.4

Proof. By Lemma 3.1, $\mathrm{T} = \mathrm{D}_\sigma - \nabla G$ is Lipschitz and pseudo-contractive. Therefore, according to Ishikawa's Theorem (Ishikawa, 1974), PnPI-GD converges strongly in $Fix(\mathrm{D}_\sigma - \nabla G)$. □

H. Proof of Theorem 3.5

Proof. By Lemma B.1, since ∇G is γ-cocoercive, the proximal operator $\mathrm{Prox}_{\frac{G}{\beta}}$ is $\frac{1}{2\gamma+2}$-averaged. Since $k < \frac{2\gamma+1}{2\gamma+2} = 1 - \frac{1}{2\gamma+2}$, by Lemma 3.2, $\mathrm{T} = \mathrm{D}_\sigma \circ \mathrm{Prox}_{\frac{G}{\beta}}$ is l-strictly pseudo-contractive, where

$$0 \le l = \frac{k(1-\frac{1}{2\gamma+2})}{(1-\frac{1}{2\gamma+2}) - k\frac{1}{2\gamma+2}} = \frac{k(2\gamma+1)}{2\gamma+1-k} < 1. \tag{54}$$

D_σ is k-strictly pseudo-contractive, and thus, D_σ is $\frac{1+k}{1-k}$-Lipschitz. Since $\mathrm{Prox}_{\frac{G}{\beta}}$ is 1-Lipschitz, $\mathrm{T} = \mathrm{D}_\sigma \circ \mathrm{Prox}_{\frac{G}{\beta}}$ is also Lipschitz. Therefore, T is Lipschitz and pseudo-contractive. According to Ishikawa's Theorem (Ishikawa, 1974), when $Fix(\mathrm{D}_\sigma \circ \mathrm{Prox}_{\frac{G}{\beta}}) \neq \emptyset$, PnPI-HQS converges strongly in $Fix(\mathrm{D}_\sigma \circ \mathrm{Prox}_{\frac{G}{\beta}})$. □

I. Proof of Theorem 3.6

Proof. ∇G is γ-cocoercive. After some derivations, we have that for any $x, y \in V$,

$$\langle x-y, \nabla G(x)-\nabla G(y)\rangle - \gamma\|\nabla G(x)-\nabla G(y)\|^2 \iff \|(2\gamma\nabla G-\mathrm{I})(x)-(2\gamma\nabla G-\mathrm{I})(y)\|^2 \leq \|x-y\|^2. \quad (55)$$

It means that $2\gamma\nabla G-\mathrm{I}$ is non-expansive. For $0 \leq \lambda \leq 2\gamma$,

$$\mathrm{I}-\lambda\nabla G = \left(1-\frac{\lambda}{2\gamma}\right)\mathrm{I}+\frac{\lambda}{2\gamma}(\mathrm{I}-2\gamma\nabla G). \quad (56)$$

Therefore, $\mathrm{I}-\lambda\nabla G$ is $\frac{\lambda}{2\gamma}$-averaged. By Lemma B.1, when D_σ is k-strictly pseudo-contractive, $\mathrm{D}_\sigma \circ (\mathrm{I}-\lambda\nabla G)$ is l-strictly pseudo-contractive, where

$$0 \leq l = \frac{k(1-\frac{\lambda}{2\gamma})}{(1-\frac{\lambda}{2\gamma})-k\frac{\lambda}{2\gamma}} = \frac{k(2\gamma-\lambda)}{2\gamma-\lambda-k\lambda} < 1, \quad (57)$$

if $k \leq 1-\frac{\lambda}{2\gamma}$. Since D_σ is k-strictly pseudo-contractive, it is $\frac{1+k}{1-k}$-Lipschitz. Then $\mathrm{T} = \mathrm{D}_\sigma \circ (\mathrm{I}-\lambda\nabla G)$ is Lipschitz. Under the assumption that $Fix(\mathrm{T}) \neq \emptyset$, Ishikawa's Theorem (Ishikawa, 1974) guarantees the strong convergence of PnPI-FBS in $Fix(\mathrm{T})$.

□

J. Denoising results on Set12

In Table 5, we provide denoising performance on Set12. In Table 6, we validate the assumptions by calculating the maximum spectral norms on Set12.

Table 5. Average denoising PSNR performance of different denoisers on Set12 dataset, for various noise levels σ.

σ	15	25	40
FFDNet	32.08	29.99	27.90
DnCNN	32.88	30.46	28.26
DRUNet	**33.08**	**30.80**	**28.76**
MMO	31.36	29.06	27.00
NE-DRUNet	31.68	29.57	27.18
Prox-DRUNet	31.71	29.04	26.45
SPC-DRUNet	32.90	30.59	28.44
PC-DRUNet	**33.01**	**30.69**	**28.66**

Table 6. Maximal values of different norms on Set12 dataset for various noise levels σ.

σ	15	25	40	Norm
DRUNet	2.134	3.936	6.070	$\|\mathrm{J}\|_*$
DRUNet	1.628	2.715	3.515	$\|\frac{1}{2}\mathrm{I}+\frac{1}{2}\mathrm{J}\|_*$
DRUNet	5.436	2.143	2.219	$\|(\mathrm{S}-2\mathrm{I})^{-1}\mathrm{S}\|_*$
PC-DRUNet ($r=10^{-3}$)	0.990	0.993	0.998	$\|\frac{1}{2}\mathrm{I}+\frac{1}{2}\mathrm{J}\|_*$
PC-DRUNet ($r=10^{-4}$)	1.001	1.296	1.465	$\|\frac{1}{2}\mathrm{I}+\frac{1}{2}\mathrm{J}\|_*$
SPC-DRUNet ($r=10^{-3}$)	0.994	0.999	0.998	$\|(\mathrm{S}-2\mathrm{I})^{-1}\mathrm{S}\|_*$
SPC-DRUNet ($r=10^{-4}$)	0.999	1.150	1.083	$\|(\mathrm{S}-2\mathrm{I})^{-1}\mathrm{S}\|_*$

K. Blur kernels

The blurring kernels used in the experiments are shown in Fig. 4.

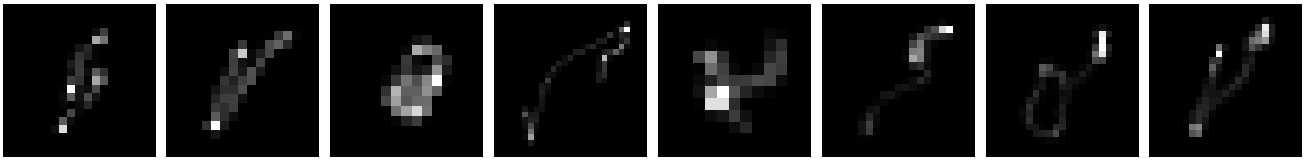


Figure 4. Eight kernels from (Levin et al., 2009).

Table 7. Average deblurring PSNR and SSIM performance by different methods on CBSD68 dataset with Levin's 8 kernels with $\sigma = 12.75$.

	kernel1	kernel2	kernel3	kernel4	kernel5	kernel6	kernel7	kernel8	Average
MMO-FBS	25.97	25.73	26.41	25.52	27.50	27.04	26.59	26.07	26.35
	0.6985	0.6871	0.7151	0.6723	0.7642	0.747	0.7282	0.7027	0.7100
NE-PGD	26.26	25.93	26.58	25.70	27.69	27.33	26.77	26.33	26.58
	0.7147	0.699	0.7252	0.6845	0.7749	0.7628	0.7413	0.7192	0.7277
Prox-DRS	26.30	25.92	26.50	25.87	27.81	26.92	27.23	26.56	26.64
	0.694	0.6829	0.7167	0.6807	0.7749	0.7451	0.7598	0.7215	0.7200
DPIR	27.43	27.21	27.61	26.98	28.57	28.34	27.75	27.29	27.65
	0.7649	0.755	0.7696	0.7434	0.8092	0.8018	0.7814	0.7651	0.7738
PnPI-GD	26.14	25.95	26.37	25.65	27.36	27.22	26.50	26.10	26.41
	0.6717	0.6729	0.7027	0.6581	0.7373	0.7240	0.7099	0.6932	0.6962
PnPI-HQS	**27.55**	**27.33**	**27.69**	**27.10**	**28.68**	**28.50**	**27.77**	**27.35**	**27.75**
	0.7719	**0.7641**	**0.7762**	**0.7531**	**0.8142**	**0.8091**	**0.7822**	**0.7668**	**0.7797**
PnPI-FBS	26.78	26.23	26.57	26.01	27.87	27.76	26.90	26.54	26.83
	0.7376	0.7170	0.7329	0.7042	0.7889	0.7836	0.7583	0.7381	0.7451

Table 8. Average deblurring PSNR and SSIM performance by different methods on CBSD68 dataset with Levin's 8 kernels with $\sigma = 17.85$.

	kernel1	kernel2	kernel3	kernel4	kernel5	kernel6	kernel7	kernel8	Average
MMO-FBS	25.56	25.39	26.01	25.16	26.76	26.36	25.96	25.59	25.72
	0.6856	0.6764	0.7007	0.6625	0.7351	0.7216	0.7041	0.6873	0.7000
NE-PGD	25.67	25.39	26.03	25.17	26.89	26.56	26.09	25.72	25.94
	0.6859	0.6724	0.6994	0.6590	0.7400	0.7292	0.7099	0.6908	0.6983
Prox-DRS	25.85	25.62	26.31	25.34	27.13	25.90	25.61	26.14	25.99
	0.6926	0.6704	0.7131	0.6646	0.7361	0.6875	0.6773	0.6922	0.6900
DPIR	26.41	26.26	26.78	26.04	27.60	27.28	26.85	**26.42**	26.75
	0.7226	0.7122	0.7326	0.7000	0.7751	0.7631	0.7452	**0.7452**	0.7293
PnPI-GD	25.28	25.21	25.69	24.90	26.46	26.34	25.70	25.27	25.61
	0.6329	0.6404	0.6760	0.6254	0.7036	0.6887	0.6782	0.6609	0.6633
PnPI-HQS	**26.53**	**26.31**	**26.84**	**26.11**	**27.69**	**27.45**	**26.90**	26.33	**26.77**
	0.7302	**0.7209**	**0.7450**	**0.7159**	**0.7796**	**0.7710**	**0.7536**	0.7252	**0.7386**
PnPI-FBS	25.82	25.52	25.96	25.32	26.98	26.86	26.29	25.88	25.97
	0.6838	0.6819	0.7047	0.6701	0.7517	0.7425	0.7256	0.7045	0.7081

L. Deblurring results on CBSD68

For a detailed PSNR and SSIM values on each kernel, see Tables 7-8.

M. Deblurring results on Set12

An overall deblurring results on Set12 by different methods are summarized in Table 9.

Table 9. Average deblurring PSNR and SSIM performance by different methods on Set12 dataset with Levin's 8 kernels with $\sigma = 12.75$ and 17.85.

	$\sigma = 12.75$		$\sigma = 17.85$	
	PSNR	SSIM	PSNR	SSIM
MMO-FBS	26.10	0.7542	25.33	0.7242
NE-PGD	26.37	0.7651	25.55	0.7363
Prox-DRS	26.55	0.7557	25.51	0.7179
DPIR	27.64	0.7987	26.48	0.7632
PnPI-GD	26.91	0.7409	25.92	0.7114
PnPI-HQS	**27.69**	**0.8020**	**26.57**	**0.7719**
PnPI-FBS	26.79	0.7790	26.01	0.7530

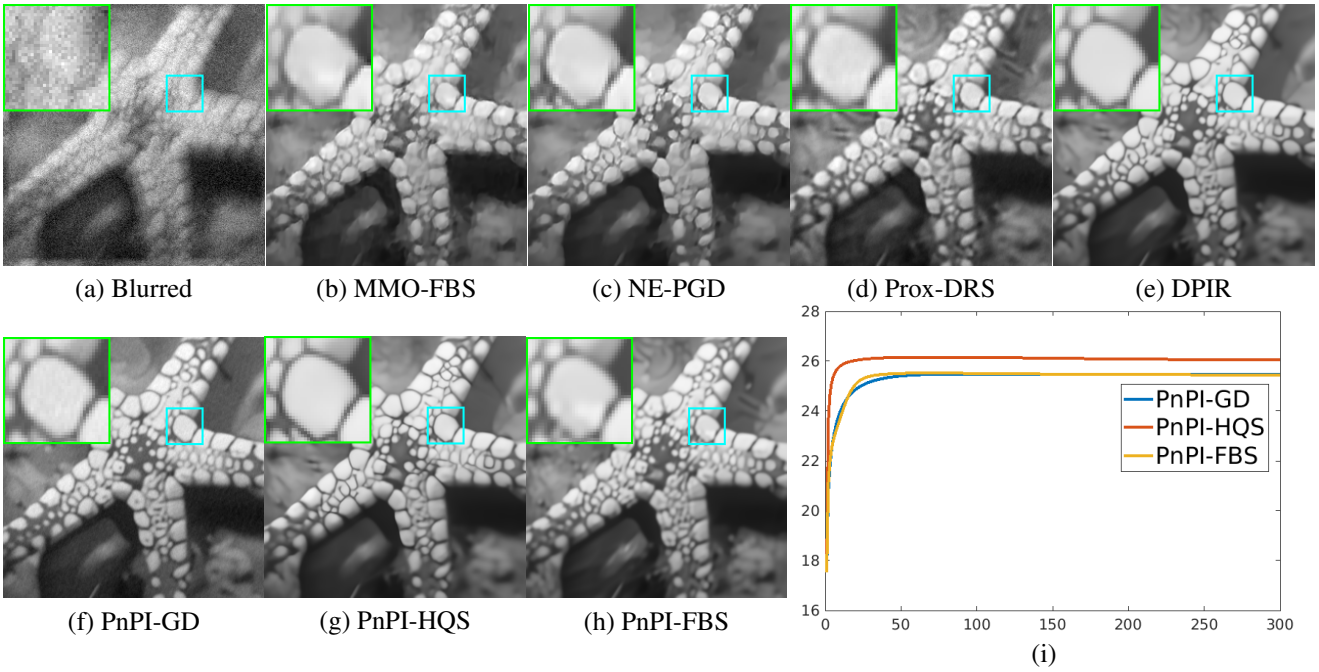


Figure 5. Results by different methods when recovering the image 'Starfish' from kernel 8 and Gaussian noise with $\sigma = 12.75$. (a) Blurred. (b) MMO-FBS, PSNR=24.64dB. (c) NE-PGD, PSNR=24.98dB. (d) Prox-DRS, PSNR=24.92dB. (e) DPIR, PSNR=25.92dB. (f) PnPI-GD, PSNR=25.49dB. (g) PnPI-HQS, PSNR=26.03dB. (h) PnPI-FBS, PSNR=25.46dB. (i) PSNR curves by PnPI-GD, PnPI-HQS, and PnPI-FBS.

In Fig. 5, it can be seen from the enlarged parts that, PnPI-HQS provides the best visual result with sharp edges. Compared with MMO-FBS, and NE-PGD in Figs. 5 (b)-(d), results by Prox-DRS, PnPI-GD and PnPI-FBS have clearer structures.

Additional results are shown in Fig. 6 with kernel 4 and $\sigma = 17.85$. PnPI-GD and PnPI-FBS provide competitive results compared with Figs. 6 (b)-(d). Result by DPIR has clear edges, but seems to have some fake structures. Among the methods, PnPI-HQS provides best visual effects with sharp edges.

A detailed PSNR and SSIM values with each kernels are listed in Tables 10-11.

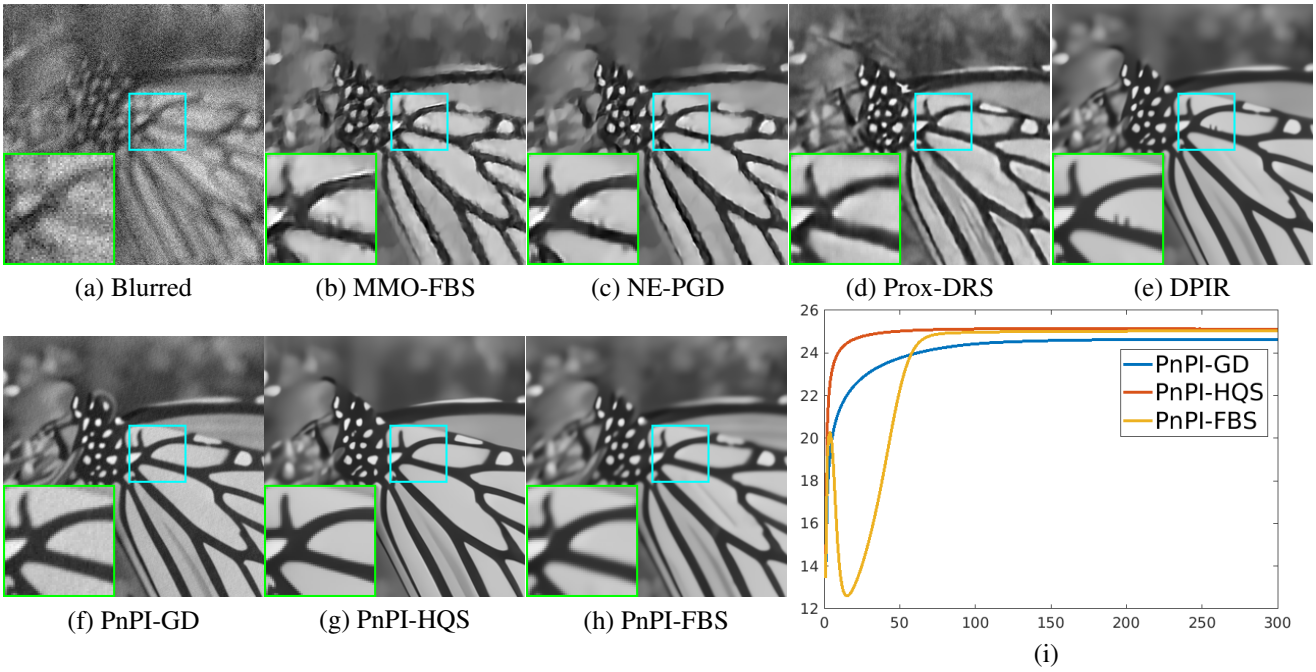


Figure 6. Results by different methods when recovering the image 'Monarch' from kernel 4 and Gaussian noise with $\sigma = 17.85$. (a) Blurred. (b) MMO-FBS, PSNR=22.83dB. (c) NE-PGD, PSNR=23.28dB. (d) Prox-DRS, PSNR=23.54dB. (e) DPIR, PSNR=24.94dB. (f) PnPI-GD, PSNR=24.58dB. (g) PnPI-HQS, PSNR=25.19dB. (h) PnPI-FBS, PSNR=25.01dB. (i) PSNR curves by PnPI-GD, PnPI-HQS, and PnPI-FBS.

Table 10. Average deblurring PSNR and SSIM performance by different methods on Set12 dataset with Levin's 8 kernels with $\sigma = 12.75$.

	kernel1	kernel2	kernel3	kernel4	kernel5	kernel6	kernel7	kernel8	Average
MMO-FBS	25.61	25.44	26.39	25.22	27.24	26.72	26.45	25.69	26.10
	0.7400	0.7363	0.7670	0.7251	0.7870	0.7706	0.7644	0.7430	0.7542
NE-PGD	25.92	25.74	26.57	25.49	27.45	27.11	26.68	26.04	26.37
	0.7516	0.7467	0.7738	0.7354	0.7974	0.7843	0.7766	0.7550	0.7651
Prox-DRS	26.23	26.02	26.78	25.76	27.69	27.31	26.85	25.74	26.55
	0.7468	0.7433	0.7682	0.7314	0.7875	0.7757	0.7649	0.7275	0.7557
DPIR	27.37	27.08	27.71	26.93	28.56	28.24	27.90	**27.31**	27.64
	0.7895	0.7824	0.8010	0.7783	0.8231	0.8137	0.8084	**0.7935**	0.7987
PnPI-GD	26.58	26.36	26.96	26.10	27.85	27.54	27.31	26.58	26.91
	0.7169	0.7249	0.7581	0.7143	0.7673	0.7526	0.7588	0.7342	0.7409
PnPI-HQS	**27.43**	**27.14**	**27.75**	**26.95**	**28.65**	**28.33**	**27.94**	27.29	**27.69**
	0.7936	**0.7858**	**0.8048**	**0.7789**	**0.8300**	**0.8188**	**0.8109**	0.7930	**0.8020**
PnPI-FBS	26.63	25.99	26.70	25.81	27.89	27.84	27.19	26.27	26.79
	0.7727	0.7552	0.7789	0.7484	0.8121	0.8049	0.7936	0.7663	0.7790

Table 11. Average deblurring PSNR and SSIM performance by different methods on Set12 dataset with Levin's 8 kernels with $\sigma = 17.85$.

	kernel1	kernel2	kernel3	kernel4	kernel5	kernel6	kernel7	kernel8	Average
MMO-FBS	24.95	24.82	25.76	24.51	26.32	25.80	25.55	24.92	25.33
	0.7137	0.7113	0.7433	0.6989	0.7514	0.7340	0.7281	0.7128	0.7242
NE-PGD	25.18	25.03	25.89	24.72	26.49	26.12	25.76	25.22	25.55
	0.7259	0.7218	0.7502	0.7101	0.7619	0.7505	0.7435	0.7266	0.7363
Prox-DRS	25.25	25.04	25.88	24.56	26.68	26.19	25.66	24.81	25.51
	0.7153	0.7101	0.7403	0.6904	0.7545	0.7413	0.7154	0.6760	0.7179
DPIR	26.17	25.97	26.66	25.73	27.33	27.03	26.74	**26.15**	26.48
	0.7528	0.7483	0.7677	0.7412	0.7877	0.7776	0.7737	0.7566	0.7632
PnPI-GD	25.61	25.46	26.13	25.10	26.83	26.54	26.18	25.54	25.92
	0.6827	0.6940	0.7338	0.6801	0.7399	0.7241	0.7270	0.7093	0.7114
PnPI-HQS	**26.32**	**26.07**	**26.81**	**25.73**	**27.59**	**27.14**	**26.82**	26.07	**26.57**
	0.7629	**0.7556**	**0.7796**	**0.7445**	**0.8037**	**0.7886**	**0.7819**	**0.7585**	**0.7719**
PnPI-FBS	25.85	25.37	26.01	25.07	27.03	26.80	26.38	25.62	26.01
	0.7451	0.7328	0.7554	0.7216	0.7842	0.7733	0.7678	0.7439	0.7530

N. Single image super-resolution results on Set12

An overall PSNR and SSIM values are summarized in Table 12. Compared with other convergent PnP methods, the proposed methods are competitive, see Table 12.

Table 12. Average super-resolution PSNR and SSIM performance by different methods on Set12 dataset with different scales and noise levels.

	s=2			s=4		
σ	0	2.55	7.65	0	2.55	7.65
MMO-FBS	27.18	26.32	25.18	25.42	25.18	24.23
	0.8247	0.7776	0.7162	0.7453	0.7341	0.6915
NE-PGD	27.17	26.45	25.23	25.51	25.26	24.30
	0.8242	0.7817	0.7315	0.7482	0.7370	0.6955
Prox-DRS	31.25	26.96	25.48	25.89	25.27	24.08
	0.9108	0.7927	0.7366	0.7810	0.7425	0.6831
DPIR	30.99	27.49	25.79	**26.56**	**25.94**	24.42
	0.8976	0.8082	0.7458	**0.7945**	**0.7648**	0.6984
PnPI-GD	27.13	26.59	25.54	25.95	25.57	24.42
	0.8179	0.7919	0.7468	0.7735	0.7539	0.7021
PnPI-HQS	**31.87**	**27.52**	**25.98**	26.29	25.72	**24.61**
	0.9128	**0.8115**	**0.7584**	0.7905	0.7573	**0.7090**
PnPI-FBS	28.96	26.50	25.59	25.99	25.66	24.49
	0.8767	0.7891	0.7473	0.7678	0.7539	0.7048

In Fig. 7, we see that, the result by PnPI-GD is smooth with clear edges, while PnPI-FBS recovers some textures. Note that in Figs. 7 (d) and (g), Prox-DRS can recover some parts of textures of the tie, but PnPI-HQS can also recover the textures on the trousers.

O. Poisson denoising on Set12

In the Poisson noise removal task, we seek to solve the inverse problem (1) with K be the identity operator and Poisson noise, that is

$$f \sim \frac{Poisson(u \times peak)}{peak}, \tag{58}$$

where $peak > 0$ determines the noise level. A large $peak$ corresponds to a low noise level. Note that in this case, the gray value interval of u is $[0, 1]$. The fidelity term is

$$G(u; f) = \mu\langle u - f \log u, \mathbf{1}\rangle. \tag{59}$$

The proximal operator $Prox_{\frac{G}{\beta}}$ can be solved according to (Kumar & Ranjan Sahay, 2019). Since $\nabla G(u) = \mathbf{1} - \frac{f}{u}$ is not cocoercive, gradient-based methods are not guaranteed to converge, such as MMO-FBS, NE-PGD, PnPI-GD, and

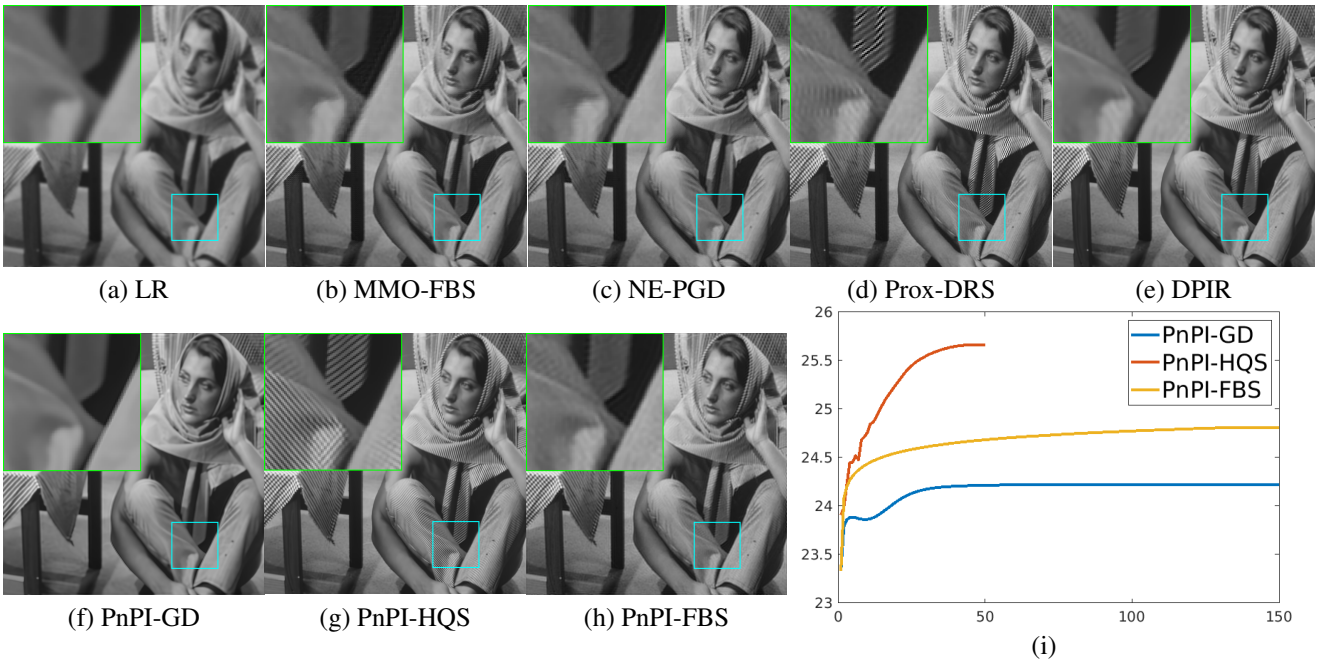


Figure 7. Super-resolution results by different methods on the image 'Barbara' with $s = 2, \sigma = 0$. (a) Low-resolution (LR). (b) MMO-FBS, PSNR=24.21dB. (c) NE-PGD, PSNR=24.24dB. (d) Prox-DRS, PSNR=24.81dB. (e) DPIR, PSNR=25.01dB. (f) PnPI-GD, PSNR=24.23dB. (g) PnPI-HQS, PSNR=25.66dB. (h) PnPI-FBS, PSNR=24.83dB. (i) PSNR curves by PnPI-GD, PnPI-HQS, and PnPI-FBS.

PnPI-FBS. However, we still compare these methods when removing Poisson noises. Although PnPI-GD and PnPI-FBS are not guaranteed to converge, we observe in experiments that both algorithms converge efficiently, see Fig. 8. We set $N = 300$, and fine tune μ, λ, β.

The overall PSNR and SSIM values are listed in Table 13. The highest value is marked in **boldface**. It can be seen that in most cases, PnPI-FBS provides the best PSNR values, while PnPI-HQS has the best SSIM values. Compared with the state-of-the-art PnP methods, the proposed PnPI-GD, PnPI-HQS, and PnPI-FBS provides competitive results.

In Fig. 8, we show the Poisson noise removal results on the image 'Lena' with $peak = 20$. In Fig. 8 (a), the enlarged part is severely degraded. The methods in Figs. 8 (b)-(e) can recover some textures. Note that in Figs. 8 (f)-(h), the proposed methods restore finer textures, with less noise residuals.

Table 13. Average Poisson denoising PSNR and SSIM performance by different methods on Set12 dataset with different peaks.

	peak=10		peak=15		peak=20	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MMO-FBS	24.75	0.7102	25.98	0.7486	26.67	0.7642
NE-PGD	25.57	0.7293	26.50	0.7634	27.20	0.7892
Prox-DRS	26.27	0.7468	26.71	0.7682	27.21	0.7834
DPIR	25.97	0.7316	27.09	0.7923	27.87	0.8180
PnPI-GD	25.90	0.7270	27.03	0.7567	27.53	0.7797
PnPI-HQS	25.82	**0.7796**	27.20	**0.8043**	28.11	**0.8212**
PnPI-FBS	**26.80**	0.7689	**27.58**	0.7827	**28.16**	0.8010

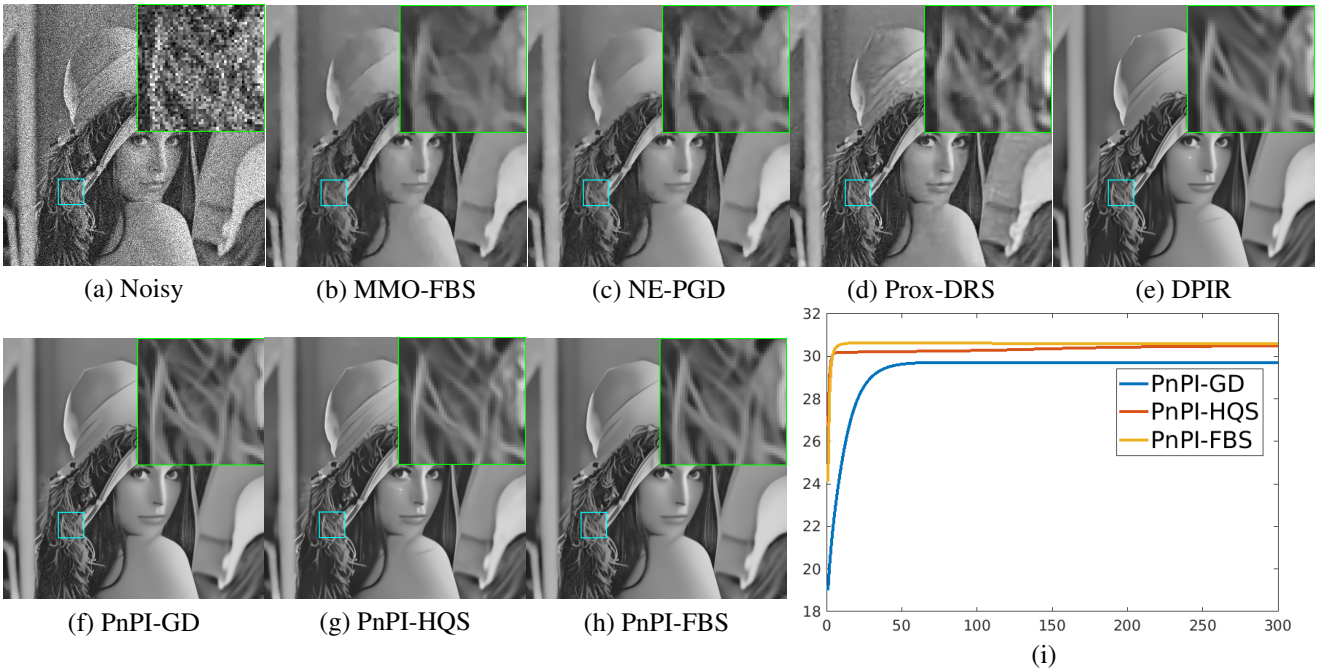


Figure 8. Results by different methods when recovering the image 'Lena' from Poisson noise (peak=20). (a) Noisy. (b) MMO-FBS, PSNR=29.39dB. (c) NE-PGD, PSNR=29.84dB. (d) Prox-DRS, PSNR=30.24dB. (e) DPIR, PSNR=30.14dB. (f) PnPI-GD, PSNR=29.73dB. (g) PnPI-HQS, PSNR=30.43dB. (h) PnPI-FBS, PSNR=30.82dB. (i) PSNR curves by PnPI-GD, PnPI-HQS, and PnPI-FBS.

P. Traffic data completion

In the spatiotemporal traffic data imputation task, we seek to solve the following tensor completion problem (Chen et al., 2024):

$$\min_{\mathcal{M}} \sum_{k=1}^{3} F\left(\mathbf{M}_{(k)}\right), \text{ s.t. } \mathcal{P}_{\Omega}\left(\mathcal{M}\right) = \mathcal{P}_{\Omega}\left(\mathcal{Y}\right) \tag{60}$$

where $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the observed tensor with missing values, $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the target spatiotemporal traffic tensor, and Ω is the index set of the observed entries in $\mathcal{Y}$, n_1, n_2, and n_3 denote the sensors, days, and time-intervals of the traffic data, respectively. $\mathbf{M}_{(k)} = \text{unfold}_k\left(\mathcal{M}\right)$ is the mode-k unfolding matrix of $\mathcal{M}$, F denotes the latent regularization term related to our pseudo-contractive denoiser, and the operator $\mathcal{P}_{\Omega} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a mask operator supported on Ω:

$$\left[\mathcal{P}_{\Omega}\left(\mathcal{Y}\right)\right]_{i_1 i_2 i_3} = \begin{cases} \mathcal{Y}_{i_1 i_2 i_3}, & \text{if } (i_1, i_2, i_3) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

By introducing a series of auxiliary matrix variables $\mathbf{Z}_k$, we transform the optimal problem Eq. (60) into the following form, and utilize the HQS method to solve it:

$$\begin{aligned} \min_{\substack{\mathcal{M}, \\ \mathbf{Z}_k, k=1,2,3}} & \sum_{k=1}^{3} \left(F\left(\mathbf{Z}_k\right) + \frac{\beta}{2} \left\| \mathbf{Z}_k - \mathbf{M}_{(k)} \right\|_F^2 \right) \\ & + \chi_{[v_{\min},\, v_{\max}]}\left(\mathcal{M}\right), \text{ s.t. } \mathcal{P}_{\Omega}\left(\mathcal{M}\right) = \mathcal{P}_{\Omega}\left(\mathcal{Y}\right). \end{aligned} \tag{61}$$

where $\chi_{[v_{\min},\, v_{\max}]}$ is an indicator function on $[v_{\min},\, v_{\max}]$, and for any $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have

$$\chi_{[v_{\min},\, v_{\max}]}\left(\mathcal{X}\right) = \sum_{i_1, i_2, i_3} \chi_{[v_{\min},\, v_{\max}]}\left(\mathcal{X}_{i_1 i_2 i_3}\right), \tag{62}$$

where

$$\chi_{[v_{\min},\, v_{\max}]}(a) = \begin{cases} 0, & \text{if } a \in [v_{\min},\, v_{\max}]\,, \\ +\infty, & \text{otherwise.} \end{cases}$$

According to Eq. (61), we have:

$$\begin{aligned} &\operatorname{Prox}_{\frac{G}{\beta}}\left(\{\mathbf{Z}_k^n\}_{k=1}^3\right) \\ &= \arg\min_{\mathcal{M}} \chi_{[v_{\min},\, v_{\max}]}\left(\mathcal{M}\right) + \frac{1}{2}\left\|\mathcal{M} - \mathcal{W}^n\right\|_F^2 \end{aligned} \tag{63}$$

on the constrain of $\mathcal{P}_\Omega\left(\mathcal{M}\right) = \mathcal{P}_\Omega\left(\mathcal{Y}\right)$, where

$$\mathcal{W}^n = \frac{1}{3}\sum_{k=1}^{3} \operatorname{fold}_k\left(\mathbf{Z}_k^n\right),$$

fold_k is the inverse operation of unfold_k, and n denotes the iteration step.

Based on the formulation of $\chi_{[v_{\min},\, v_{\max}]}$ in Eq. (62) and the conclusion from Lemma 6.26 in Section 6.4.2 of the book (Beck, 2017), we can get

$$\begin{aligned} &\left[\operatorname{Prox}_{\frac{G}{\beta}}\left(\{\mathbf{Z}_k^n\}_{k=1}^3\right)\right]_{i_1 i_2 i_3} \\ &= \begin{cases} \mathcal{Y}_{i_1 i_2 i_3}, & (i_1, i_2, i_3) \in \Omega, \\ \min\left(v_{\max},\, \max\left(\mathcal{W}_{i_1 i_2 i_3}^n,\, v_{\min}\right)\right), & (i_1, i_2, i_3) \notin \Omega. \end{cases} \end{aligned} \tag{64}$$

Therefore we can get PnPI-HQS (Algorithm 6) to solve Eq. (60):

Algorithm 6 PnPI-HQS for solving Eq. (60)

Input: Observation spatiotemporal traffic data $\mathcal{Y}$, masked index set Ω, $\{\alpha_n\}$, $\{\beta_n\}$, $v_{\min}$, $v_{\max}$, N, ϵ, D_σ;
Initialize $\mathbf{Z}_k^0$ as zeros, $k = 1, 2, 3$, set $\mathcal{M}^0$ such that $\mathcal{M}_\Omega^0 = \mathcal{Y}_\Omega$, $\mathcal{M}_{\Omega^\perp}^0 = \operatorname{mean}\left(\mathcal{Y}_\Omega\right)$, $n = 0$;
for $n = 0$ to $N - 1$ **do**
  for $k = 1$ to 3 **do**
    $\mathbf{Z}_k^n = \left(1 - \beta_n\right)\mathbf{M}_{(k)}^n + \beta_n\, \mathrm{D}_\sigma\left(\mathbf{M}_{(k)}^n\right)$;
  end for
  $\mathcal{Y}^n = \operatorname{Prox}_{\frac{G}{\beta}}\left(\{\mathbf{Z}_k^n\}_{k=1}^3\right)$;
  for $k = 1$ to 3 **do**
    $\widehat{\mathbf{M}}_k^n = \left(1 - \alpha_n\right)\mathbf{M}_{(k)}^n + \alpha_n\, \mathrm{D}_\sigma\left(\mathbf{Y}_{(k)}^n\right)$;
  end for
  $\mathcal{M}^{n+1} = \operatorname{Prox}_{\frac{G}{\beta}}\left(\left\{\widehat{\mathbf{M}}_k^n\right\}_{k=1}^3\right)$;
  Calculate $e = \frac{\left\|\mathcal{M}^{n+1} - \mathcal{M}^n\right\|_F}{\left\|\mathcal{M}^n\right\|_F}$;
  if $e < \epsilon$ **then**
    Stop iteration.
  end if
end for
Output: $\mathcal{M}^{n+1}$.

In Algorithm 6, the initialization $\mathcal{M}_\Omega^0 = \mathcal{Y}_\Omega$, $\mathcal{M}_{\Omega^\perp}^0 = \operatorname{mean}\left(\mathcal{Y}_\Omega\right)$ represents that:

$$\mathcal{M}_{i_1 i_2 i_3}^0 = \begin{cases} \mathcal{Y}_{i_1 i_2 i_3}, & (i_1,\, i_2,\, i_3) \in \Omega, \\ \operatorname{mean}\left(\mathcal{Y}_\Omega\right) = \dfrac{1}{|\Omega|}\displaystyle\sum_{ijk \in \Omega} \mathcal{Y}_{ijk}, & (i_1,\, i_2,\, i_3) \notin \Omega. \end{cases}$$

We take PeMS freeway traffic volume dataset as a concrete example to assess the imputation performance of Algorithm 6. The PeMS dataset $(228 \times 44 \times 288)$ includes the traffic volume by 228 loop detectors in District 7 of California, with a 5-minute time resolution. Within this task, we consider two classical missing scenarios: random missing (RM) and blackout missing (BM). The generation method of RM and BM can refer to (Chen et al., 2019b) and (Chen et al., 2021), respectively.

To measure the methods' performance, we manually mask a few entries to substitute for the missing cases for the missing scenarios and compare the imputed values with masked ones. The evaluation of completion performance metrics are mean absolute percentage error (MAPE) and root mean square error (RMSE):

$$\text{MAPE} = \frac{1}{c}\sum_{i=1}^{c}\frac{|y_i - \widehat{y}_i|}{|y_i|} \times 100, \ \text{RMSE} = \sqrt{\frac{1}{c}\sum_{i=1}^{c}(y_i - \widehat{y}_i)^2},$$

where y_i and $\widehat{y}_i$ are the actual value and the imputation value, respectively, and c is the total numbers of estimated values, i.e., $c = |\Omega_m|$, where $\Omega_m = \{(i,\, j,\, k) \mid y_{ijk}$ is masked and not missing in original tensor$\}$. The smaller value means better completion performance.

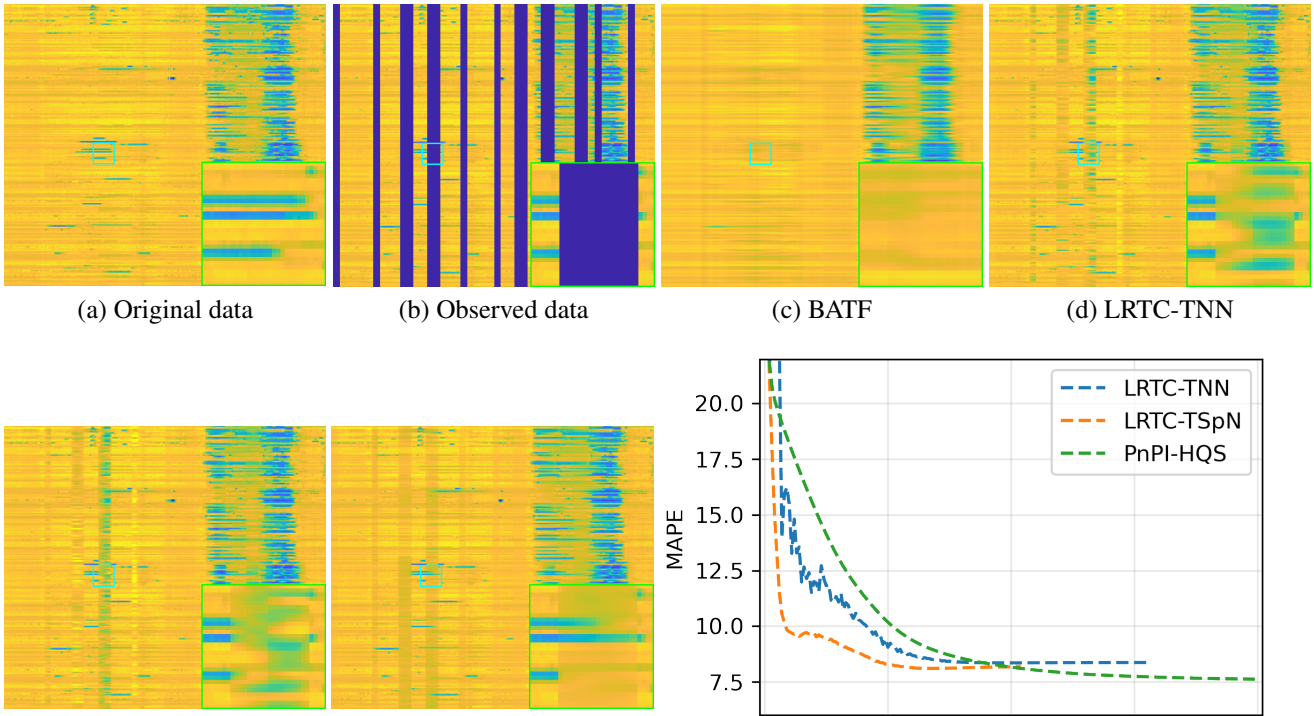
For comparison, we choose three baseline methods for spatiotemporal traffic data imputaiton, BATF (Chen et al., 2019a), LRTC-TNN (Chen et al., 2020), and LRTC-TSpN (Nie et al., 2022). Table 14 exhibits the imputation results on different missing scenarios. It is clear that the proposed Algorithm 6 outperforms other baseline methods in simple missing conditions.

Table 14. Imputation Results (MAPE / RMSE).

Dataset	Missing Pattern	BATF	LRTC-TNN	LRTC-TSpN	PnPI-HQS
PeMS	10 % RM	6.78 / 4.68	2.80 / 1.97	2.28 / 1.65	**2.10 / 1.60**
	30 % RM	6.82 / 4.68	3.45 / 2.44	2.74 / 1.96	**2.27 / 1.80**
	50 % RM	6.83 / 4.71	4.45 / 3.13	3.42 / 2.42	**2.69 / 2.21**
	70 % RM	6.96 / 4.78	5.94 / 4.15	4.63 / 3.21	**3.73 / 3.08**
	10 % BM-12	8.58 / 5.86	7.58 / 5.35	7.40 / 5.31	**6.33 / 4.88**
	30 % BM-12	8.89 / 6.07	8.37 / 5.81	8.18 / 5.76	**7.62 / 5.80**
Average		7.48 / 5.13	5.43 / 3.81	4.77 / 3.38	**4.12 / 3.23**

To further demonstrate the imputation performance, we illustrate the enlarged completion results of different methods on the box region in Fig. 9. Compared to the original data, it is difficult to achieve precise results due to the absence of values from all sensors during this continuous period. However, our method produces the most accurate results, maintaining the original data's local consistency. BATF tends to over-smooth its results, only recovering the rough structure of the original data. As for LRTC-TNN and LRTC-TSpN, the over-shrinkage issue of their proposed low-rank regularization causes their results to generate biased turning values.

Additionally, we present the convergence curve of MAPE versus iterations in Fig. 9 (g), providing a clear reflection of the convergence ability of each method. The figure shows that all methods effectively and stably decrease the objective function. Among them, our proposed method achieves the smallest MAPE value. LRTC-TNN and LRTC-TSpN can attain convergence through similar iteration steps, while LRTC-TSpN can achieve a smaller MAPE value. LRTC methods can achieve more precise imputation results for challenging missing scenarios (Chen et al., 2024), and our proposed method has the potential for future improvement.



(a) Original data (b) Observed data (c) BATF (d) LRTC-TNN

(e) LRTC-TSpN (f) PnPI-HQS

(g) Curves

Figure 9. Enlarged visualization of the imputation results. (a) Original data; (b) Observed data; (c) BATF; (d) LRTC-TNN; (e) LRTC-TSpN; (f) PnPI-HQS; (g) The MAPE variation curve of PeMS 30% BM-12.