Rebuttal for the paper #2705

Anonymous

April 2, 2025

Abstract

In this PDF, we list several major revisions for the submitted paper "Learning Cocoercive Conservative Denoisers via Helmholtz-Hodge Decomposition for Poisson Imaging Inverse Problems". The additional experiments are listed in Sections 1-5. The theoretical revisions are listed in Sections 6-7. All the theoretical revisions are marked in blue.

In experiments, we report the computation cost in seconds in section 1, ablation study on the parameter α_1 and α_2 in section 3, and visual performance under extreme conditions in section 2. Besides, we have conducted several real-world experiments, including blind deblurring in section 4, and blind denoising in section 5. These experiments show the efficiency and effectiveness of the proposed methods in real-world applications.

In theory, we corrected the original Theorem 2.2. Besides, we fortunately find that the original characterization Theorem 2.2 is not tight enough: when t is small, ∂F is also Lipschitz, see the current Theorem 6.2 for details. Since Theorem 6.2 has been revised, the convergence theorem is also strengthened, see the current Theorem 7.1 for details. Now, by Theorem 6.2 and Theorem 7.1, when $\gamma = 0.25$, $0 \le t < t_0 = \frac{1}{3}$, CoCo-ADMM converges. In every old and new experiment, we uniformly set $\gamma = 0.25$ and $t = 0.33$. Thus CoCo-ADMM is guaranteed to converge.

1 Computational time

In Table 1, we give the average computation time in seconds by different methods when deblurring a 256×256 image. It can be seen that the proposed methods are efficient enough for real-time applications.

	B-RED	Prox-DRS	SNORE	DiffPIR	DPS	CoCo-ADMM	CoCo-PEGD
Iteration	300	100	200	100	1000	50	100
Time	42.5s	13.6s	74.1s	11.8s	177s	5.24s	9.08s

Table 1: Average computation time in seconds by different methods when deblurring a 256×256 image.

2 Extreme case

In Fig. 1, we show an example when deblurring an image under severe noisy condition. It can be seen in Fig. 1 (a) that the image is severely corrupted. We compare results by different methods. It can be seen that CoCo-ADMM and CoCo-PEGD provide sharper edges. Even in this extreme case, the proposed methods are still convergent, see Figs. 1 (m)-(n).

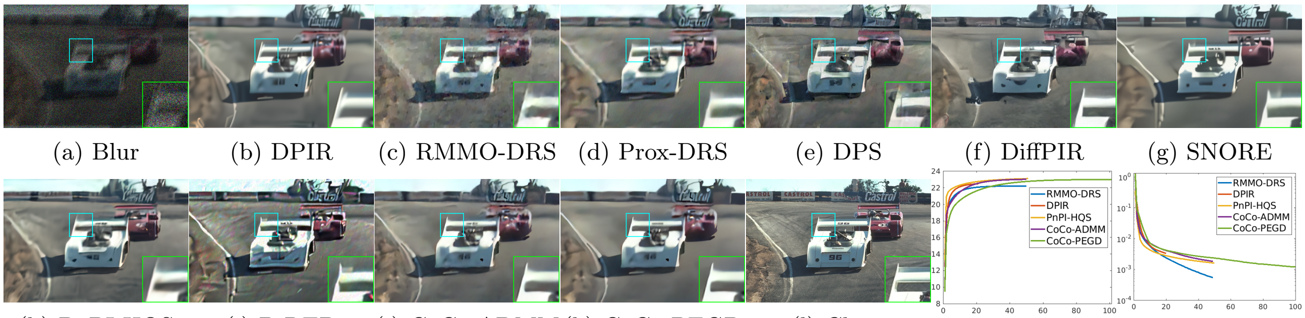
3 Ablation study on the parameters α_1 and α_2

$$Loss(\theta) = \mathbb{E}\| \mathrm{D}_\sigma(x+\xi;\theta) - x\|_1 + \alpha_1 \| \mathrm{J} - \mathrm{J}^\top \|_* + \alpha_2 \max\{\|2\gamma \mathrm{J} - \mathrm{I}\|_*, 1-\epsilon\}. \tag{1}$$

In the loss function (1), α_1 controls the penalty strength of the Hamiltonian term to encourage conservativeness, and α_2 controls the penalty strength of the spectral term to encourage cocoerciveness.

As shown in Table 2, DRUNet without regularization terms has an expansive residual part. When α_1 gets bigger, the mean symmetry error gets lower. When α_2 is smaller than 1e-2, the regularized denoiser may not be cocoercive.

Therefore, we set the penalty parameters large enough to encourage the properties we want. However, when α gets too large, the denoising performance may be sacrificed. We empirically set $\alpha_1 = 1$, $\alpha_2 =$ 1e-2 in the experiments.



(a) Blur (b) DPIR (c) RMMO-DRS (d) Prox-DRS (e) DPS (f) DiffPIR (g) SNORE

(h) PnPI-HQS (i) B-RED (j) CoCo-ADMM (k) CoCo-PEGD (l) Clean (m) PSNR (n) Relative error

Figure 1: Deconvolution results by different methods on the image '0005' from CBSD68 with kernel 8 and $p = 10$ Poisson noises. (a) Blur image. (b) DPIR, PSNR=22.49dB. (c) RMMO-DRS, PSNR=22.23dB. (d) Prox-DRS, PSNR=22.54dB. (e) DPS, PSNR=20.70dB. (f) DiffPIR, PSNR=21.66dB. (g) SNORE, PSNR=23.31dB. (h) PnPI-HQS, PSNR=23.14dB. (i) B-RED, PSNR=21.15dB. (j) CoCo-ADMM, PSNR=23.04dB. (k) CoCo-PEGD, PSNR=23.02dB. (l) Clean image. (m) PSNR curves. (n) Relative error curves. x-axis denotes the iteration number.

Table 2: Mean symmetry error $\|\mathrm{J}-\mathrm{J}^\top\|_*$ and maximal values of the norm $\|2\gamma\,\mathrm{J}-\mathrm{I}\|_*$ on CBSD68 for various noise levels σ and $\gamma = 0.50, 0.25$.

	15	25	40	Norms	α_1	α_2
DRUNet	1.7e+1	1.8e+3	3.3e+2	$\|\mathrm{J}-\mathrm{J}^\top\|_*$	0	0
0.50-CoCo-DRUNet	7.4e-7	6.3e-8	4.9e-7	$\|\mathrm{J}-\mathrm{J}^\top\|_*$	1	1e-2
0.50-CoCo-DRUNet	8.5e-2	7.0e-2	9.1e-1	$\|\mathrm{J}-\mathrm{J}^\top\|_*$	1e-2	1e-2
0.25-CoCo-DRUNet	9.6e-8	3.7e-8	1.6e-7	$\|\mathrm{J}-\mathrm{J}^\top\|_*$	1	1e-2
0.25-CoCo-DRUNet	3.5e-3	7.3e-3	6.2e-3	$\|\mathrm{J}-\mathrm{J}^\top\|_*$	1e-2	1e-2
DRUNet	3.285	4.343	6.283	$\|\mathrm{J}-\mathrm{I}\|_*$	0	0
0.50-CoCo-DRUNet	0.994	0.992	0.972	$\|\mathrm{J}-\mathrm{I}\|_*$	1	1e-2
0.50-CoCo-DRUNet	1.011	1.244	1.500	$\|\mathrm{J}-\mathrm{I}\|_*$	1	1e-3
0.25-CoCo-DRUNet	0.986	0.969	0.982	$\|0.5\,\mathrm{J}-\mathrm{I}\|_*$	1	1e-2
0.25-CoCo-DRUNet	0.992	1.010	1.211	$\|0.5\,\mathrm{J}-\mathrm{I}\|_*$	1	1e-3

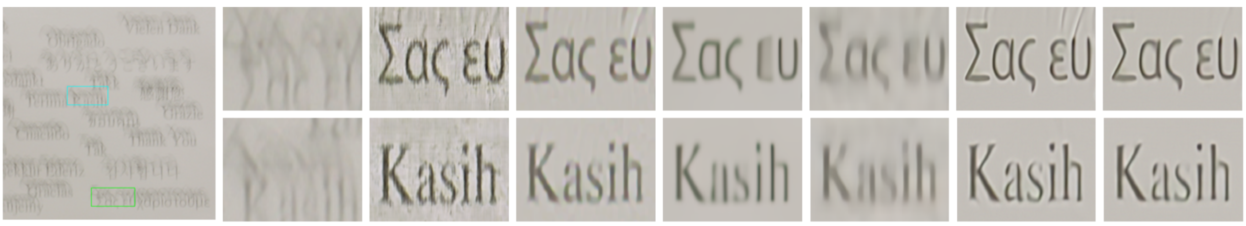


Figure 2: Blind decovolution results by different methods.

	INR	Lv et al.	DPIR	RMMO	CoCo-ADMM	CoCo-PEGD
Average	0.2990	0.3444	0.3629	0.2664	0.5461	**0.5474**

Table 3: Average CLIP-IQA results across all deblurred images

4 Real Deblur Results

In this section, we evaluate the performance of our proposed methods on real-world blind deblurring. We select 18 images from the dataset by Lai et al. [5], which captures challenging real-world blurry scenes. Four state-of-the-art methods are benchmarked: Deblur-INR [14], Lv et al. [7], DPIR [15], and RMMO [10]. Deblur-INR is a self-supervised method that jointly estimates clean images and blur kernels, while Lv et al. [7] applies total variation regularization to refine latent image estimation in a blind deblurring framework. In contrast, DPIR, RMMO, and our proposed CoCo-ADMM/CoCo-PEGD require pre-estimated blur kernels as input. To ensure fairness, all kernel-dependent methods utilize blur kernels generated by the blind deblurring framework of Liu et al. [7]. Visual comparisons in Fig. 2 demonstrate that CoCo-ADMM and CoCo-PEGD produce sharper textures and fewer artifacts than baseline methods.

For quantitative evaluation, we employ the CLIP-IQA metric [12], a non-reference image quality assessment tool that leverages the pretrained vision-language representation of CLIP (Contrastive Language–Image Pre-training) to evaluate perceptual quality without requiring pristine reference images. The average CLIP-IQA scores across all deblurred images are summarized in Table 3. Despite relying on the same estimated kernels as DPIR and RMMO, our methods achieve superior performance, highlighting their robustness to kernel estimation errors and enhanced restoration capability.

5 Real Denoise Results

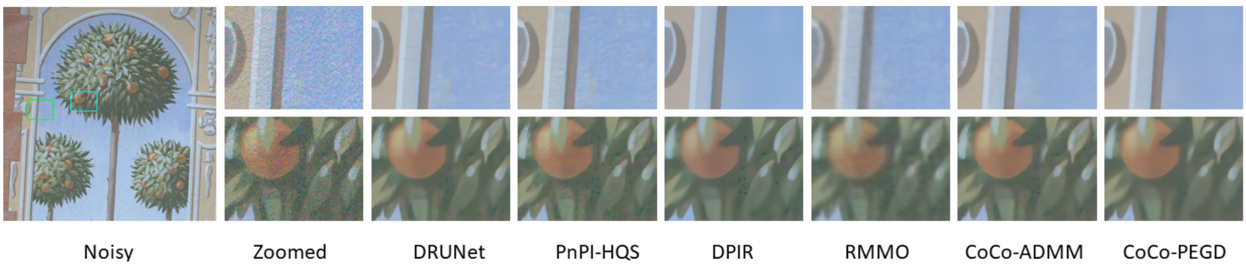


Figure 3: Blind noise removal results by different methods.

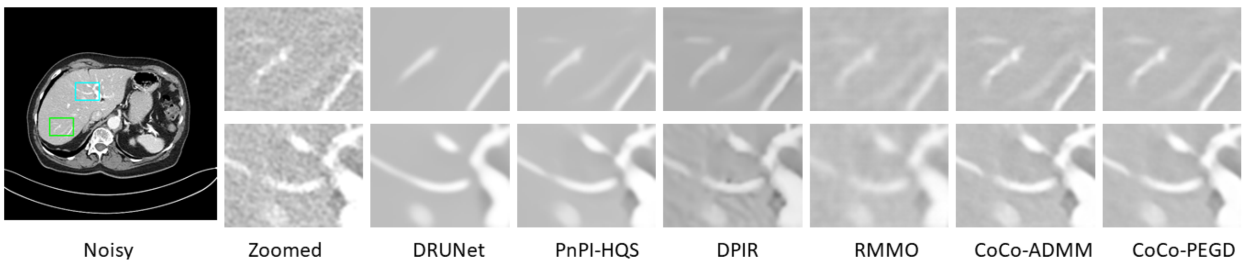


Figure 4: Blind noise suppression results in CT images by different methods.

In this section, we evaluate the performance of our proposed denoising framework on two benchmark datasets that capture real-world noise characteristics. For medical imaging, we employ the Low-Dose CT dataset from Mayo Clinic [8], which provides paired low-dose CT scans with Poisson noise.

For natural images, we utilize the DND dataset [9], comprising 50 real noisy photographs corrupted by a mixture of Poisson and Gaussian noise. Following the standard evaluation protocol, perceptual quality is quantified using the CLIP-IQA metric [12].

To ensure a fair comparison, we benchmark our proposed methods (CoCo-ADMM and CoCo-PEGD) against five representative techniques: DRUNet[15], DPIR [15] , RMMO [10] , and PnPI-HQS [13]. Quantitative results in Table 4 demonstrate that our proposed methods achieve state-of-the-art performance across both imaging domains.

	DRUNet	DPIR	RMMO	PnPI-HQS	CoCo-ADMM	CoCo-PEGD
Color images	0.4464	0.4509	0.4561	0.4586	**0.4596**	0.4593
CT	0.4154	0.4258	0.4398	0.4011	0.4707	**0.4730**

Table 4: Average CLIP-IQA results for blind image denoising.

6 Better Characterization Theorem for CoCo Denoisers

We have already proved the following theorem in the original manuscript. Here we state this theorem, and omit the proof, because the proof was correct already.

Theorem 6.1. *Let* $\mathrm{D}_\sigma \in \mathcal{C}^1[V]$. $\beta = \frac{1}{\sigma^2}$. D_σ *satisfies that:*
- D_σ *is conservative;*
- D_σ *is* γ*-cocoercive with* $\gamma \in (0, \infty)$.

Then, there exists a function $F : V \to \bar{\mathbb{R}}$, *such that* F *is* r*-weakly convex, where* $r = \beta(1-\gamma)$, *and that* $\mathrm{D}_\sigma(x) \in \operatorname{Prox}_{\frac{F}{\beta}}(x), \forall x \in V$.

The following Theorem 6.2, that was the Theorem 2.2 originally, has been corrected and strengthened. Specifically speaking, now by Theorem 6.2, for all $\gamma \in (0, 1)$, $t \in [0, 1)$, ∂F is L-Lipschitz.

Theorem 6.2. *Let* $\mathrm{D}_\sigma \in \mathcal{C}^1[V]$. $\beta = \frac{1}{\sigma^2}$. D_σ *satisfies that:*
- D_σ *is conservative;*
- D_σ *is* γ*-cocoercive with* $\gamma \in (0, 1)$.

Let $\mathrm{D}_\sigma^t = t\,\mathrm{D}_\sigma + (1-t)\,\mathrm{I}$, $t \in [0, 1)$. *It holds that:*
- *there exists a* r*-weakly convex function* $F : V \to \bar{\mathbb{R}}$, $r(t) = \beta\frac{t-\gamma t}{t+\gamma-\gamma t}$, *such that* $\mathrm{D}_\sigma^t(x) \in \operatorname{Prox}_{\frac{F}{\beta}}(x), \forall x \in V$;
- ∂F *is* L*-Lipschitz, where*

$$L(t) = \begin{cases} \beta\frac{t}{1-t} \geq r(t), & \text{if } t \geq \frac{1-2\gamma}{2-2\gamma} \text{ and } t \in [0,1) \text{ (Case 1)}; \\ r(t) = \beta\frac{t-\gamma t}{t+\gamma-\gamma t}, & \text{if } t \leq \frac{1-2\gamma}{2-2\gamma} \text{ and } t \in [0,1) \text{ (Case 2)}. \end{cases} \tag{2}$$

Proof. Since D_σ is γ-cocoercive with $\gamma > 0$, D_σ^t is naturally $\frac{\gamma}{t+(1-t)\gamma}$-cocoercive: $\forall x, y \in V$, we have

$$\begin{aligned} \langle \mathrm{D}_\sigma(x) - \mathrm{D}_\sigma(y), x-y \rangle &\geq \gamma \| \mathrm{D}_\sigma(x) - \mathrm{D}_\sigma(y)\|^2 \\ \langle \frac{1}{t}(\mathrm{D}_\sigma^t - (1-t)\,\mathrm{I}) \circ (x) - \frac{1}{t}(\mathrm{D}_\sigma^t - (1-t)\,\mathrm{I}) \circ (y), x-y \rangle &\geq \gamma \| \tfrac{1}{t}(\mathrm{D}_\sigma^t - (1-t)\,\mathrm{I}) \circ (x) - \tfrac{1}{t}(\mathrm{D}_\sigma^t - (1-t)\,\mathrm{I}) \circ (y)\|^2 \\ t\langle \mathrm{D}_\sigma^t(x) - \mathrm{D}_\sigma^t(y), x-y \rangle - t(1-t)\|x-y\|^2 &\geq \gamma \| \mathrm{D}_\sigma^t(x) - \mathrm{D}_\sigma^t(y) - (1-t)(x-y)\|^2. \end{aligned} \tag{3}$$

Denote $a = \mathrm{D}_\sigma^t(x) - \mathrm{D}_\sigma^t(y)$, $b = x - y$ for convenience. Then we have

$$\begin{aligned} t\langle a, b\rangle - t(1-t)\|b\|^2 &\geq \gamma\|a\|^2 - 2\gamma(1-t)\langle a, b\rangle + \gamma(1-t)^2\|b\|^2, \\ (t + 2\gamma(1-t))\langle a, b\rangle &\geq \gamma\|a\|^2 + (t(1-t) + \gamma(1-t)^2)\|b\|^2. \end{aligned} \tag{4}$$

Now note that

$$\langle a, b\rangle = t\langle \mathrm{D}_\sigma(x) - \mathrm{D}_\sigma(y), x-y\rangle + (1-t)\|x-y\|^2. \tag{5}$$

Since D_σ is γ-cocoercive with $\gamma > 0$, we know that D_σ is $\frac{1}{\gamma}$-Lipschitz. Therefore, by Cauchy-Schwarz inequality,

$$\langle \mathrm{D}_\sigma(x) - \mathrm{D}_\sigma(y), x-y\rangle \leq \frac{1}{\gamma}\|x-y\|^2. \tag{6}$$

That is,

$$\begin{aligned} \langle a, b\rangle &\leq \left(\tfrac{t}{\gamma} + 1 - t\right)\|b\|^2, \\ \gamma(1-t)\langle a, b\rangle &\leq t(1-t)\|b\|^2 + \gamma(1-t)^2\|b\|^2. \end{aligned} \tag{7}$$

By substituting (7) into (4), we have

$$\begin{aligned} (t + \gamma(1-t))\langle a, b\rangle &\geq \gamma\|a\|^2, \\ \langle a, b\rangle &\geq \frac{\gamma}{t+\gamma(1-t)}\|a\|^2, \end{aligned} \tag{8}$$

which means that D_σ^t is $\frac{\gamma}{t+\gamma(1-t)}$-cocoercive. Since D_σ is conservative, D_σ^t is also conservative. By Theorem 6.1, we know that there exists a r-weakly convex function $F : V \to \bar{\mathbb{R}}$, where

$$r = r(t) = \beta\left(1 - \frac{\gamma}{t+\gamma(1-t)}\right) = \beta\frac{t-\gamma t}{t+\gamma-\gamma t},$$

such that $\mathrm{D}_\sigma^t \in \operatorname{Prox}_{\frac{F}{\beta}}$.

Now we prove that ∂F is L-Lipschitz. We consider two cases seperately.
Case 1: $t \geq \frac{1-2\gamma}{2-2\gamma}$ and $t \in [0, 1)$.

In this case, we need to prove that ∂F is $L(t)=\beta\frac{t}{1-t}$ Lipschitz. Given $\forall x, y \in V$, choose arbitrary u, v, such that, $u=\mathrm{D}_\sigma^t(x) \in (\mathrm{I}+\frac{1}{\beta}\partial F)^{-1}(x), v=\mathrm{D}_\sigma^t(y) \in (\mathrm{I}+\frac{1}{\beta}\partial F)^{-1}(y)$, then

$$\beta(x-u) \in \partial F(u), \beta(y-v) \in \partial F(v). \tag{9}$$

Note that $a=u-v, b=x-y$. In order to prove ∂F is $L(t)$-Lipschitz, we need to prove

$$\begin{aligned}
\|\beta(x-u)-\beta(y-v)\|^2 &\leq L^2(t)\|u-v\|^2=\beta^2\frac{t^2}{(1-t)^2}\|u-v\|^2, \\
\|\beta(b-a)\|^2 &\leq \beta^2\frac{t^2}{(1-t)^2}\|a\|^2, \\
\|b-a\|^2 &\leq \frac{t^2}{(1-t)^2}\|a\|^2.
\end{aligned} \tag{10}$$

Since $a=u-v=\mathrm{D}_\sigma^t(x)-\mathrm{D}_\sigma^t(y)=t(\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y))+(1-t)(x-y)$, $b=x-y$, we have

$$a-b=t(\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y))-t(x-y). \tag{11}$$

Now we only need to prove

$$\begin{aligned}
& t^2\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2-2t^2\langle\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y), x-y\rangle+t^2\|x-y\|^2 \\
\leq\ & \frac{t^2}{(1-t)^2}\left(t^2\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2+2t(1-t)\langle\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y), x-y\rangle+(1-t)^2\|x-y\|^2\right),
\end{aligned} \tag{12}$$

which is equivalent to prove

$$\frac{1-2t}{2-2t}\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2 \leq \langle\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y), x-y\rangle. \tag{13}$$

Since $t \geq \frac{1-2\gamma}{2-2\gamma}$, we have that

$$\frac{1-2t}{2-2t} \leq \frac{1-2\frac{1-2\gamma}{2-2\gamma}}{2-2\frac{1-2\gamma}{2-2\gamma}}=\frac{2-2\gamma-2(1-2\gamma)}{2(2-2\gamma)-2(1-2\gamma)}=\frac{2\gamma}{2}=\gamma. \tag{14}$$

We already have

$$\gamma\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2 \leq \langle\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y), x-y\rangle. \tag{15}$$

Thus,

$$\frac{1-2t}{2-2t}\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2 \leq \gamma\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2 \leq \langle\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y), x-y\rangle. \tag{16}$$

Case 2 (new added): $t \leq \frac{1-2\gamma}{2-2\gamma}$ **and** $t \in [0,1)$**.**

In this case, we need to prove that ∂F is $L(t)=r(t)=\beta\frac{t-t\gamma}{t+\gamma-t\gamma}$ Lipschitz. Similarly in (10), we need to prove

$$\begin{aligned}
\|\beta(x-u)-\beta(y-v)\|^2 &\leq L^2(t)\|u-v\|^2=\beta^2\frac{t^2(1-\gamma)^2}{(t+\gamma-t\gamma)^2}\|u-v\|^2, \\
\|b-a\|^2 &\leq \frac{t^2(1-\gamma)^2}{(t+\gamma-t\gamma)^2}\|a\|^2.
\end{aligned} \tag{17}$$

Since $a=u-v=\mathrm{D}_\sigma^t(x)-\mathrm{D}_\sigma^t(y)=t(\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y))+(1-t)(x-y)$, $b=x-y$, we have

$$a-b=t(\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y))-t(x-y). \tag{18}$$

Now we only need to prove

$$\begin{aligned}
& t^2\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2-2t^2\langle\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y), x-y\rangle+t^2\|x-y\|^2 \\
\leq\ & \frac{t^2(1-\gamma)^2}{(t+\gamma-t\gamma)^2}\left(t^2\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2+2t(1-t)\langle\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y), x-y\rangle+(1-t)^2\|x-y\|^2\right).
\end{aligned} \tag{19}$$

When $t=0$, it holds naturally. When $t \in (0,1)$, and $t \leq \frac{1-2\gamma}{2-2\gamma}$, it is equivalent to prove

$$\begin{aligned}
& \|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2-2\langle\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y), x-y\rangle+\|x-y\|^2 \\
\leq\ & \frac{(1-\gamma)^2}{(t+\gamma-t\gamma)^2}\left(t^2\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2+2t(1-t)\langle\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y), x-y\rangle+(1-t)^2\|x-y\|^2\right),
\end{aligned} \tag{20}$$

which is equivalent to prove

$$\begin{array}{rl} & (t+\gamma-t\gamma)^2\left[\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2-2\langle \mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y),x-y\rangle+\|x-y\|^2\right] \\ \leq & (1-\gamma)^2\left[t^2\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2+2t(1-t)\langle \mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y),x-y\rangle+(1-t)^2\|x-y\|^2\right], \end{array} \tag{21}$$

that is to prove

$$\begin{array}{l} \left[(t+\gamma-t\gamma)^2-t^2(1-\gamma)^2\right]\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2-\left[2(t+\gamma-t\gamma)^2+2t(1-t)(1-\gamma)^2\right]\langle \mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y),x-y\rangle \\ \leq \left[(1-t)^2(1-\gamma)^2-(t+\gamma-t\gamma)^2\right]\|x-y\|^2. \end{array} \tag{22}$$

Now we check each coefficient, and estimate each term carefully. For the coefficient of $\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2$, we have

$$(t+\gamma-t\gamma)^2-t^2(1-\gamma)^2=2\gamma(1-\gamma)t+\gamma^2>0 \tag{23}$$

For the coefficient of $\langle \mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y),x-y\rangle$, it is obviously non-positive. Besides, we have that

$$\begin{array}{rl} & -\left[2(t+\gamma-t\gamma)^2+2t(1-t)(1-\gamma)^2\right] \\ = & -2\left[(1-\gamma)^2t^2+2\gamma(1-\gamma)t+\gamma^2+(t-t^2)(1-\gamma)^2\right] \\ = & -(2-2\gamma^2)t-2\gamma^2. \end{array} \tag{24}$$

Since D_σ is γ-cocoercive, we have that

$$\left[-(2-2\gamma^2)t-2\gamma^2\right]\langle \mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y),x-y\rangle \leq \left[-\gamma(2-2\gamma^2)t-2\gamma^3\right]\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2. \tag{25}$$

For the coefficient of $\|x-y\|^2$, note that when $\gamma\in(0,1)$, $t\in[0,1)$, and $t\leq\frac{1-2\gamma}{2-2\gamma}$, we have

$$(1-t)^2(1-\gamma)^2-(t+\gamma-t\gamma)^2=(2\gamma-2)t-2\gamma+1\geq(2\gamma-2)\frac{1-2\gamma}{2-2\gamma}-2\gamma+1\geq 0. \tag{26}$$

Combining (23)-(26), to prove (22), we only need to prove that

$$\begin{array}{rrl} & \left[2\gamma(1-\gamma)t+\gamma^2-\gamma(2-2\gamma^2)t-2\gamma^3\right]\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2 & \leq\left[(2\gamma-2)t-2\gamma+1\right]\|x-y\|^2 \\ \iff & \left[(2\gamma^3-2\gamma^2)t+\gamma^2-2\gamma^3\right]\|\mathrm{D}_\sigma(x)-\mathrm{D}_\sigma(y)\|^2 & \leq\left[(2\gamma-2)t-2\gamma+1\right]\|x-y\|^2. \end{array} \tag{27}$$

Since D_σ is $\frac{1}{\gamma}$-Lipschitz, we only need to prove that

$$\begin{array}{rrl} & \frac{1}{\gamma^2}\left[(2\gamma^3-2\gamma^2)t+\gamma^2-2\gamma^3\right] & \leq(2\gamma-2)t-2\gamma+1 \\ \iff & (2\gamma-2)t+1-2\gamma & \leq(2\gamma-2)t-2\gamma+1. \end{array} \tag{28}$$

This completes the proof. □

7 Better Convergence Theorem for CoCo-ADMM

The restoration model is

$$\hat{u}\in\arg\min_{u\in V}F(u)+G(u;f),\quad G(u;f)=\lambda\langle\mathbf{1},Ku-f\log Ku\rangle. \tag{29}$$

CoCo-ADMM takes the form of:

$$\begin{array}{ll} u^{k+1} & =\mathrm{Prox}_{\frac{G}{\beta}}(v^k-b^k), \\ v^{k+1} & =\mathrm{D}_\sigma^t(u^{k+1}+b^k), \\ b^{k+1} & =b^k+u^{k+1}-v^{k+1}, \end{array} \tag{30}$$

where D_σ^t is defined in Theorem 6.2, and $\beta=\frac{1}{\sigma^2}$. The PnP-ADMM algorithm in (30) with a γ-CoCo denoiser is referred to as γ-CoCo-ADMM, or CoCo-ADMM for short.

When the denoiser $\mathrm{D}_\sigma\in\mathcal{C}^1[V]$ is a CoCo denoiser satisfying the conditions in Theorem 6.2, and F verifies the Kurdyka-Lojasiewicz (KL) property [1, 3], the global convergence of CoCo-ADMM in (30) can be established as follows.

Since Theorem 6.2 has been revised, the convergence theorem is also revised. In special, now by Theorem 7.1, for all $\gamma\in(0,1)$ and $t\in[0,t_0)$, CoCo-ADMM converges. The major revision lies in part 1, where we prove that when $t<t_0$, an important value $\frac{\beta}{2}-\frac{r}{2}-\frac{L^2}{\beta}$ is positive.

Theorem 7.1. *Let $F : V \to \bar{\mathbb{R}}$ be a coercive weakly convex KL function in Theorem 6.2 such that $\mathrm{D}^t_\sigma \in \mathrm{Prox}_{\frac{F}{\beta}}$. $G : V \to \bar{\mathbb{R}}$ is lower semi-continuous and convex. $\gamma \in (0,1)$. $t \in [0, t_0)$, where $t_0 = t_0(\gamma)$ is the positive root of the equation*

$$(2-2\gamma)t^3 + \gamma t^2 + 2\gamma t - \gamma = 0. \tag{31}$$

Then, the sequence $\{(u^k, v^k, b^k)\}$ generated by (30) converges globally to a point (u^*, v^*, b^*), and that $u^* = v^*$ is a stationary point of the model (29).

We will make use of the Lyapunov function L_β for (30) according to [6, 11, 4]:

$$L_\beta(u, v, b) = F(v) + G(u; f) + \beta\langle b, u - v\rangle + \frac{\beta}{2}\|u - v\|^2. \tag{32}$$

We will first prove in part 1 that an important value for $L_\beta(u, v, b)$ is positive whenever $t \in (0, t_0)$, where t_0 is the positive root of the characteristic equation in (31). Then, we will prove in part 2 that L_β is non-increasing with the iteration number k. Finally, we will prove in part 3 that CoCo-ADMM iteration in (30) converges globally to a stationary point of (29).

Proof. Let $h(t) = (2-2\gamma)t^3 + \gamma t^2 + 2\gamma t - \gamma$, where $\gamma \in (0, 1)$. Note that h is obviously smooth, and $h(0) = -\gamma < 0$, $h(\infty) = \infty$. Also note that when $t > 0$,

$$h'(t) = (2-2\gamma)t^2 + 2\gamma t + 2\gamma > 0. \tag{33}$$

Therefore, there exists a unique $t_0 > 0$, such that $h(t_0) = 0$, $h(t) > 0$ if $t > t_0$, and $h(t) < 0$ if $t \in [0, t_0)$.

Part 1:

We consider a characteristic value for $L_\beta(u, v, b)$ in (32): $\frac{\beta}{2} - \frac{r}{2} - \frac{L^2}{\beta}$.

By Theorem 6.2, if $t \in [0, 1)$ and $t \geq \frac{1-2\gamma}{2-2\gamma}$, we have that $r = \beta\frac{t-\gamma t}{t+\gamma-t\gamma}$ and $L = \beta\frac{t}{1-t}$. Thus we have

$$\begin{aligned}
& \frac{\beta}{2} - \frac{r}{2} - \frac{L^2}{\beta} = \frac{\beta}{2} - \frac{\beta(t-\gamma t)}{2(t+\gamma-t\gamma)} - \frac{\beta t^2}{(1-t)^2} \\
= \;& \frac{\beta}{2}\left(1 - \frac{t-\gamma t}{t+\gamma-t\gamma} - \frac{2t^2}{(1-t)^2}\right) \\
= \;& \frac{\beta}{2(t+\gamma-t\gamma)(1-t)^2}\left(\gamma(1-t)^2 - 2t^2(t+\gamma-t\gamma)\right) \\
= \;& -\frac{\beta}{2(t+\gamma-t\gamma)(1-t)^2}\left((2-2\gamma)t^3 + \gamma t^2 + 2\gamma t - \gamma\right).
\end{aligned} \tag{34}$$

When $0 < t < t_0$, where t_0 is the positive root of the characteristic equation in (31), $\frac{\beta}{2} - \frac{r}{2} - \frac{L^2}{\beta} > 0$ holds.

If $t \in [0, 1)$ and $t \leq \frac{1-2\gamma}{2-2\gamma}$, we have that $L = r = \beta\frac{t-\gamma t}{t+\gamma-t\gamma}$. Note that in this case, $L = r \leq \beta\frac{t}{1-t}$. Thus,

$$\begin{aligned}
& \frac{\beta}{2} - \frac{r}{2} - \frac{L^2}{\beta} \\
\geq \;& \frac{\beta}{2} - \frac{\beta(t-\gamma t)}{2(t+\gamma-t\gamma)} - \frac{\beta t^2}{(1-t)^2} \\
= \;& -\frac{\beta}{2(t+\gamma-t\gamma)(1-t)^2}\left((2-2\gamma)t^3 + \gamma t^2 + 2\gamma t - \gamma\right).
\end{aligned} \tag{35}$$

Therefore, we also have that when $0 < t < t_0$, $\frac{\beta}{2} - \frac{r}{2} - \frac{L^2}{\beta} > 0$ holds.

Part 2:

Now we prove that $L_\beta(u^k, v^k, b^k)$ is non-increasing. Before that, we show two important formulas. For v^{k+1}, by the first-order optimal condition, we know that

$$\beta b^{k+1} = -\beta(v^{k+1} - u^{k+1} - b^k) \in \partial F(v^{k+1}). \tag{36}$$

Similarly, for u^{k+1}, we have

$$-\beta(u^{k+1} - v^k + b^k) \in \partial G(u^{k+1}; f). \tag{37}$$

In order to prove that $L_\beta(u^k, v^k, b^k)$ is non-increasing, we decompose $L_\beta(u^k, v^k, b^k) - L_\beta(u^{k+1}, v^{k+1}, b^{k+1})$ into two parts:

$$\begin{aligned}
& L_\beta(u^k, v^k, b^k) - L_\beta(u^{k+1}, v^{k+1}, b^{k+1}) \\
= \;& L_\beta(u^k, v^k, b^k) - L_\beta(u^{k+1}, v^k, b^k) + L_\beta(u^{k+1}, v^k, b^k) - L_\beta(u^{k+1}, v^{k+1}, b^{k+1}),
\end{aligned} \tag{38}$$

and estimate them separately as follows. By the convexity of G and the iteration form in (30), we have

$$\begin{aligned}
&L_\beta(u^k,v^k,b^k)-L_\beta(u^{k+1},v^k,b^k)\\
=\;& G(u^k;f)-G(u^{k+1};f)+\beta\langle b^k,u^k-u^{k+1}\rangle+\frac{\beta}{2}\|u^k-v^k\|^2-\frac{\beta}{2}\|u^{k+1}-v^k\|^2\\
=\;& G(u^k;f)-G(u^{k+1};f)-\langle-\beta(u^{k+1}-v^k+b^k),u^k-u^{k+1}\rangle+\beta\langle v^k-u^{k+1},u^k-u^{k+1}\rangle\\
&+\frac{\beta}{2}\|u^k-v^k\|^2-\frac{\beta}{2}\|u^{k+1}-v^k\|^2\\
\geq\;& 0+\beta\langle v^k-u^{k+1},u^k-u^{k+1}\rangle+\frac{\beta}{2}\langle u^k-u^{k+1},u^k+u^{k+1}-2v^k\rangle\\
=\;& \beta\langle u^k-u^{k+1},v^k-u^{k+1}-v^k+\frac{u^k+u^{k+1}}{2}\rangle\\
=\;& \frac{\beta}{2}\|u^k-u^{k+1}\|^2.
\end{aligned}\tag{39}$$

By the r-weakly convexity of F, $\forall x,y\in V,\ f_y\in\partial F(y)$, we have:

$$F(x)-F(y)\geq\langle f_y,x-y\rangle-\frac{r}{2}\|x-y\|^2.\tag{40}$$

By the L-Lipschitz property of ∂F, $\forall x,y\in V,\ f_y\in\partial F(y)$, we have:

$$F(x)-F(y)\leq\langle f_y,x-y\rangle+\frac{L}{2}\|x-y\|^2.\tag{41}$$

Combining (40) and (41), we can obtain that

$$\begin{aligned}
&L_\beta(u^{k+1},v^k,b^k)-L_\beta(u^{k+1},v^{k+1},b^{k+1})\\
=\;& F(v^k)-F(v^{k+1})+\beta\langle b^k,u^{k+1}-v^k\rangle-\beta\langle b^{k+1},u^{k+1}-v^{k+1}\rangle\\
&+\frac{\beta}{2}\|u^{k+1}-v^k\|^2-\frac{\beta}{2}\|u^{k+1}-v^{k+1}\|^2\\
=\;& F(v^k)-F(v^{k+1})+\beta\langle b^k,u^{k+1}-v^k\rangle-\beta\langle b^k,u^{k+1}-v^{k+1}\rangle-\beta\langle u^{k+1}-v^{k+1},u^{k+1}-v^{k+1}\rangle\\
&+\frac{\beta}{2}\|v^k-v^{k+1}\|^2+\beta\langle u^{k+1}-v^{k+1},v^{k+1}-v^k\rangle\\
=\;& F(v^k)-F(v^{k+1})+\beta\langle b^k,v^{k+1}-v^k\rangle-\beta\|u^{k+1}-v^{k+1}\|^2\\
&+\frac{\beta}{2}\|v^k-v^{k+1}\|^2+\beta\langle u^{k+1}-v^{k+1},v^{k+1}-v^k\rangle\\
=\;& F(v^k)-F(v^{k+1})-\langle\beta b^k,v^{k+1}-v^k\rangle-\beta\|u^{k+1}-v^{k+1}\|^2+\frac{\beta}{2}\|v^k-v^{k+1}\|^2\\
&+\beta\langle u^{k+1}-v^{k+1},v^{k+1}-v^k\rangle\\
\geq\;& -\frac{r}{2}\|v^k-v^{k+1}\|^2-\beta\|u^{k+1}-v^{k+1}\|^2+\frac{\beta}{2}\|v^k-v^{k+1}\|^2\\
=\;& \left(\frac{\beta}{2}-\frac{r}{2}\right)\|v^k-v^{k+1}\|^2-\beta\|b^k-b^{k+1}\|^2\\
\geq\;& \left(\frac{\beta}{2}-\frac{r}{2}\right)\|v^k-v^{k+1}\|^2-\frac{L^2}{\beta}\|v^k-v^{k+1}\|^2\\
=\;& \left(\frac{\beta}{2}-\frac{r}{2}-\frac{L^2}{\beta}\right)\|v^k-v^{k+1}\|^2.
\end{aligned}\tag{42}$$

Note that the second '=' comes from the cosine rule, the first '$\geq$' follows from the r-weakly convexity of F as in (40), and the second '$\geq$' results from the L-Lipschitz of ∂F as in (41).

Combining (39) and (42), we get

$$L_\beta(u^k,v^k,b^k)-L_\beta(u^{k+1},v^{k+1},b^{k+1})\geq\frac{\beta}{2}\|u^k-u^{k+1}\|^2+\left(\frac{\beta}{2}-\frac{r}{2}-\frac{L^2}{\beta}\right)\|v^k-v^{k+1}\|^2\geq 0,\tag{43}$$

that is, $L_\beta(u^k,v^k,b^k)$ is non-increasing.

Now we prove that $\{(u^k,v^k,b^k)\}$ is bounded. Note that F and G are coercive on V. As a result,

$$F(u^k)+G(u^k;f)>+\infty.\tag{44}$$

Along with $\beta b^k\in\partial F(v^k)$ and the property that ∂F is L-Lipschitz, we arrive at

$$\begin{aligned}
L_\beta(u^k,v^k,b^k)\;&=F(v^k)+G(u^k;f)+\beta\langle b^k,u^k-v^k\rangle+\frac{\beta}{2}\|u^k-v^k\|^2\\
&\geq F(u^k)+G(u^k;f)-\frac{L}{2}\|u^k-v^k\|^2+\frac{\beta}{2}\|u^k-v^k\|^2.
\end{aligned}\tag{45}$$

Note that $L = \frac{\beta t}{1-t}$, and that $t < 0.5$. Thus, $L < \beta$. Therefore,

$$\begin{aligned} & F(u^k) + G(u^k; f) - \frac{L}{2}\|u^k - v^k\|^2 + \frac{\beta}{2}\|u^k - v^k\|^2 \\ = \quad & F(u^k) + G(u^k; f) + \frac{1}{2}(\beta - L)\|u^k - v^k\|^2 \geq -\infty. \end{aligned} \tag{46}$$

Since $F(u) + G(u; f)$ is coercive on V, u^k, v^k, b^k are bounded.

Part 3:

Define q^{k+1} as follows:

$$q^{k+1} = [\beta(b^{k+1} - b^k + v^k - v^{k+1}), \beta(v^{k+1} - u^{k+1}), \beta(b^{k+1} - b^k)]. \tag{47}$$

Define $\partial L_\beta(u, v, b) = [\partial_u L_\beta, \partial_v L_\beta, \partial_b L_\beta]$. By the formulas (36)-(37), we know that

$$q^{k+1} \in \partial L_\beta(u^{k+1}, v^{k+1}, b^{k+1}). \tag{48}$$

Note that

$$\|\beta(b^{k+1} - b^k + v^k - v^{k+1})\| \leq \beta\|b^k - b^{k+1}\| + \beta\|v^k - v^{k+1}\| \leq L\|v^k - v^{k+1}\| + \beta\|v^k - v^{k+1}\|, \tag{49}$$

and that

$$\|\beta(v^{k+1} - u^{k+1})\| = \beta\|b^k - b^{k+1}\| \leq L\|v^k - v^{k+1}\|, \tag{50}$$

we arrive at

$$\|q^{k+1}\| \leq C\|v^k - v^{k+1}\|, \tag{51}$$

where

$$C = 3L + \beta. \tag{52}$$

Now we can finally prove Theorem 7.1. By Part 2, $\{(u^k, v^k, b^k)\}$ is bounded. So there is a sub-sequence $\{(u^{n_k}, v^{n_k}, b^{n_k})\}$ such that $(u^{n_k}, v^{n_k}, b^{n_k}) \to (u^*, v^*, b^*)$, when $n \to +\infty$. Since $L_\beta(u^k, v^k, b^k)$ is lower bounded and non-increasing, we have that $\|u^k - u^{k+1}\|, \|v^k - v^{k+1}\| \to 0$ as $k \to +\infty$. Besides, since $q^k \in \partial L_\beta(u^k, v^k, b^k)$ and $\|q^k\| \leq C\|v^k - v^{k+1}\|$, we know that $\|q^k\| \to 0$, $\|q^{n_k}\| \to 0$. Thus, $0 \in \partial L_\beta(u^*, v^*, b^*)$, and (u^*, v^*, b^*) is a stationary point of L_β.

Since F, G is KL, we conclude that L_β is also KL. Then, by the proof of Theorem 2.9 in [2], $\{(u^{n_k}, v^{n_k}, b^{n_k})\}$ converges globally to (u^*, v^*, b^*).

Since (u^*, v^*, b^*) is a stationary point of L_β, and as a result, $q^* = 0$, that is

$$q^* = [0, \beta(v^* - u^*), 0] = 0. \tag{53}$$

Therefore,$u^* = v^*$. By CoCo-ADMM iteration in (30), we know that

$$\begin{aligned} u^* &= \mathrm{Prox}_{\frac{G}{\beta}}(u^* - b^*), \\ u^* &= \mathrm{D}_\sigma^t(u^* + b^*) \in \mathrm{Prox}_{\frac{G}{\beta}}(u^* + b^*). \end{aligned} \tag{54}$$

Equivalently,

$$-\beta b^* \in \partial G(u^*; f), \beta b^* \in \partial F(u^*), \tag{55}$$

$$0 \in \partial F(u^*) + \partial G(u^*; f). \tag{56}$$

Therefore, u^* is a stationary point of (29). $\square$

References

[1] Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality. *Mathematics of operations research*, 35(2):438–457, 2010.

[2] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods. *Mathematical Programming*, 137(1):91–129, 2013.

[3] Pierre Frankel, Guillaume Garrigos, and Juan Peypouquet. Splitting methods with variable metric for kurdyka–łojasiewicz functions and general convergence rates. *Journal of Optimization Theory and Applications*, 165:874–900, 2015.

[4] Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. In *International Conference on Machine Learning*, pages 9483–9505. PMLR, 2022.

[5] Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. A comparative study for single image blind deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1709, 2016.

[6] Guoyin Li and Ting Kei Pong. Douglas–rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. *Mathematical programming*, 159:371–401, 2016.

[7] Jun Liu, Ming Yan, and Tieyong Zeng. Surface-aware blind image deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3):1041–1055, 2019.

[8] Taylor R et al. Moen. Low-dose ct image and projection dataset. *Medical Physics*, 48(2):902–911, 2021.

[9] Tobias Plötz and Stefan Roth. Benchmarking denoising algorithms with real photographs. *CVPR*, 2017.

[10] Matthieu Terris, Audrey Repetti, Jean-Christophe Pesquet, and Yves Wiaux. Building firmly nonexpansive convolutional neural networks. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8658–8662. IEEE, 2020.

[11] Andreas Themelis and Panagiotis Patrinos. Douglas–rachford splitting and admm for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181, 2020.

[12] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 2555–2563, 2023.

[13] Deliang Wei, Peng Chen, and Fang Li. Learning pseudo-contractive denoisers for inverse problems. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 52500–52524. PMLR, 21–27 Jul 2024.

[14] Tianjing Zhang, Yuhui Quan, and Hui Ji. Cross-scale self-supervised blind image deblurring via implicit neural representation. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[15] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1219–1229, 2023.