Supplementary Material

The mathematical analysis presented in this supplement builds on two distinct lines of work: (a) optimization-based characterization of the MMSE denoisers [23, 66, 82]; (b) analysis of incremental optimization algorithms [74, 83–85]. Our results are also related to two recent papers on PnP, namely the work on BC-RED in [21] and on PnP-ISTA in [23]. Our results can be viewed as an extension of [21] to nonconvex data fidelity terms and expansive denoisers. They can also be viewed as an extension of [23] to block-coordinate updates and possibly inexact MMSE denoisers.

The structure of this supplementary document is as follows. In Section A, we prove the convergence of BC-PnP under the deterministic sequential update rule. In Section B, we prove the convergence of BC-PnP under the random i.i.d. update rule. In Section C, we provide technical lemmas useful for the proofs of the main theorems. In Section D, we provide background material useful for our theoretical analysis. In Section E, we provide additional simulations omitted from the main paper due to space.

A Proof of Theorem 1

Theorem. *Run BC-PnP under Assumptions 1-5 using the sequential block selection and the step* $0 < \gamma < 1/L_{\mathsf{max}}$. *Then, we have*

$$\frac{1}{t}\sum_{i=1}^{t} \|\nabla f(\boldsymbol{x}^{ib})\|_2^2 \leq \frac{C_1}{t}(f(\boldsymbol{x}^0) - f^*) + C_2\bar{\varepsilon}_{tb}^2,$$

where $C_1 > 0$ *and* $C_2 > 0$ *are iteration independent constants. If additionally the sequence of error terms* $\{\varepsilon_i\}_{i\geq 1}$ *is square-summable, we have that* $\nabla f(\boldsymbol{x}^{tb}) \to \mathbf{0}$ *as* $t \to 0$.

Proof. The block update $i \in \{1, \cdots, b\}$ of the sequential BC-PnP using the inexact and exact denoisers can be expressed

$$\boldsymbol{x}_i^{(k-1)b+i} = \begin{cases} \boldsymbol{x}_j^{(k-1)b+i-1} & \text{if } j \neq i \\ \mathsf{D}_{\sigma_j}\left(\boldsymbol{x}_j^{(k-1)b+i-1} - \gamma\nabla_j g\left(\boldsymbol{x}^{(k-1)b+i-1}\right)\right) & \text{if } j = i \end{cases}. \tag{14}$$

and

$$\boldsymbol{z}_i^{(k-1)b+i} = \begin{cases} \boldsymbol{x}_j^{(k-1)b+i-1} & \text{if } j \neq i \\ \mathsf{D}^*_{\sigma_j}\left(\boldsymbol{x}_j^{(k-1)b+i-1} - \gamma\nabla_j g\left(\boldsymbol{x}^{(k-1)b+i-1}\right)\right) & \text{if } j = i \end{cases}. \tag{15}$$

We introduce two variables

$$\boldsymbol{v}^k \coloneqq \boldsymbol{x}^{kb} = (\boldsymbol{x}_1^{(k-1)b+1}, \cdots, \boldsymbol{x}_b^{kb}) \quad \text{and} \quad \boldsymbol{u}^k \coloneqq \boldsymbol{z}^{kb} = (\boldsymbol{z}_1^{(k-1)b+1}, \cdots, \boldsymbol{z}_b^{kb})$$

Since h_i is smooth for any $\boldsymbol{z}_i \in \mathsf{Im}(\mathsf{D}^*_{\sigma_i})$, the optimality conditions for each denoiser imply

$$\nabla_i g\left(\boldsymbol{x}^{(k-1)b+i-1}\right) + \frac{1}{\gamma}\left(\boldsymbol{z}_i^{(k-1)b+i} - \boldsymbol{x}_i^{(k-1)b+i-1}\right) + \nabla h_i\left(\boldsymbol{z}_i^{(k-1)b+i}\right) = \mathbf{0}, \tag{16}$$

for each $i \in \{1, \cdots, b\}$ and any $k \geq 1$. Since we have

$$\boldsymbol{v}^{k-1} = \boldsymbol{x}^{(k-1)b} = \left(\boldsymbol{x}_1^{(k-2)b+1}, \cdots, \boldsymbol{x}_b^{(k-1)b}\right) = \left(\boldsymbol{x}_1^{(k-1)b}, \cdots, \boldsymbol{x}_b^{kb-1}\right).$$

we can re-write (16) as

$$\frac{1}{\gamma}\left(\boldsymbol{v}^{k-1} - \boldsymbol{u}^k\right) = \begin{bmatrix} \nabla_1 g\left(\boldsymbol{x}^{(k-1)b}\right) + \nabla h_1\left(\boldsymbol{z}_1^{(k-1)b+1}\right) \\ \vdots \\ \nabla_b g\left(\boldsymbol{x}^{kb-1}\right) + \nabla h_b\left(\boldsymbol{z}_b^{kb}\right) \end{bmatrix}.$$

From the smoothness of h_i for $\boldsymbol{z}_i \in \mathsf{Im}(\mathsf{D}^*_{\sigma_i})$, we thus have that

$$\nabla f(\boldsymbol{u}^k) = \begin{bmatrix} \nabla_1 g(\boldsymbol{u}^k) + \nabla h_1\left(\boldsymbol{z}_1^{(k-1)b+1}\right) \\ \vdots \\ \nabla_b g(\boldsymbol{u}^k) + \nabla h_b\left(\boldsymbol{z}_b^{kb}\right) \end{bmatrix} = \frac{1}{\gamma}(\boldsymbol{v}^{k-1} - \boldsymbol{u}^k) + \begin{bmatrix} \nabla_1 g(\boldsymbol{u}^k) - \nabla_1 g(\boldsymbol{x}^{(k-1)b}) \\ \vdots \\ \nabla_b g(\boldsymbol{u}^k) - \nabla_b g(\boldsymbol{x}^{kb-1}). \end{bmatrix}$$

From the smoothness of g and the sequential nature of updates, we can obtain the following bound

$$\left\| \begin{bmatrix} \nabla_1 g(\boldsymbol{u}^k) - \nabla_1 g(\boldsymbol{x}^{(k-1)b}) \\ \vdots \\ \nabla_b g(\boldsymbol{u}^k) - \nabla_b g(\boldsymbol{x}^{kb-1}). \end{bmatrix} \right\|_2 \leq \sum_{i=1}^{b} \|\nabla_i g(\boldsymbol{u}^k) - \nabla_i g(\boldsymbol{x}^{(k-1)b-1+i})\|_2$$

$$\leq L \sum_{i=1}^{b} \|\boldsymbol{u}^k - \boldsymbol{x}^{(k-1)b-1+i}\|_2 \leq bL \left(\|\boldsymbol{u}^k - \boldsymbol{v}^k\|_2 + \|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2 \right),$$

where for the last inequality we used the triangular inequality. By combining the last two equations and using the step-size $\gamma = 1/(\alpha L_{\mathsf{max}})$, we get

$$\begin{aligned} \|\nabla f(\boldsymbol{u}^k)\|_2 &\leq \alpha L_{\mathsf{max}} \|\boldsymbol{u}^k - \boldsymbol{v}^{k-1}\|_2 + bL \|\boldsymbol{u}^k - \boldsymbol{v}^k\|_2 + bL \|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2 \\ &\leq (\alpha L_{\mathsf{max}} + bL) \|\boldsymbol{u}^k - \boldsymbol{v}^k\|_2 + (\alpha L_{\mathsf{max}} + bL) \|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2 \\ &\leq (\alpha L_{\mathsf{max}} + bL) \|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2 + (\alpha L_{\mathsf{max}} + bL) \sum_{i=1}^{b} \varepsilon_{(k-1)b+i}. \end{aligned}$$

By using this bound, we can get the following bound for the iterate of BC-PnP

$$\begin{aligned} \|\nabla f(\boldsymbol{v}^k)\|_2 &\leq \|\nabla f(\boldsymbol{u}^k)\|_2 + \|\nabla f(\boldsymbol{v}^k) - \nabla f(\boldsymbol{u}^k)\|_2 \\ &\leq \|\nabla f(\boldsymbol{u}^k)\|_2 + (L + M_{\mathsf{max}}) \|\boldsymbol{v}^k - \boldsymbol{u}^k\|_2 \\ &\leq \|\nabla f(\boldsymbol{u}^k)\|_2 + (L + M_{\mathsf{max}}) \sum_{i=1}^{b} \varepsilon_{(k-1)b+i} \\ &\leq A_1 \|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2 + A_2 \sum_{i=1}^{b} \varepsilon_{(k-1)b+i}. \end{aligned}$$

with $A_1 \coloneqq (\alpha L_{\mathsf{max}} + bL)$ and $A_2 \coloneqq (\alpha L_{\mathsf{max}} + bL + L + M_{\mathsf{max}})$, where we first used the triangular inequality and then Lemma 3. By squaring both sides and using $(a+b)^2 \leq 2a^2 + 2b^2$

$$\begin{aligned} \|\nabla f(\boldsymbol{v}^k)\|_2^2 &\leq 2A_1^2 \|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2^2 + 2A_2^2 \left[\sum_{i=1}^{b} \varepsilon_{(k-1)b+i} \right]^2 \\ &\leq 2A_1^2 \|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2^2 + 2bA_2^2 \sum_{i=1}^{b} \varepsilon_{(k-1)b+i}^2. \end{aligned}$$

By combining this inequality with Lemma 1, we get

$$\|\nabla f(\boldsymbol{v}^k)\|_2^2 \leq B_1 (f(\boldsymbol{v}^{k-1}) - f(\boldsymbol{v}^k)) + B_2 \sum_{i=1}^{b} \varepsilon_{(k-1)b+i}^2$$

with $B_1 \coloneqq 4A_1^2/((\alpha - 1)L_{\mathsf{max}})$ and $B_2 \coloneqq 2bA_2^2 + \lambda A_1^2$, where λ is given in Lemma 1. By averaging both sides of the bound over $t \geq 1$, we get the desired result

$$\min_{1 \leq k \leq t} \|\nabla f(\boldsymbol{v}^k)\|_2^2 \leq \frac{1}{t} \sum_{k=1}^{t} \|\nabla f(\boldsymbol{v}^k)\|_2^2 \leq \frac{C_1}{t} (f(\boldsymbol{x}^0) - f^*) + C_2 \left[\frac{1}{tb} \sum_{k=1}^{tb} \varepsilon_k^2 \right]$$

where $C_1 \coloneqq B_1$ and $C_2 \coloneqq bB_2$. □

B Proof of Theorem 2

Theorem. *Run BC-PnP under Assumptions 1-5 using the random i.i.d. block selection and the step* $0 < \gamma < 1/L_{\mathsf{max}}$. *Then, we have*

$$\mathbb{E} \left[\frac{1}{t} \sum_{k=1}^{t} \|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2 \right] \leq \frac{D_1}{t} (f(\boldsymbol{x}^0) - f^*) + D_2 \bar{\varepsilon}_t^2,$$

where $D_1 > 0$ *and* $D_2 > 0$ *are iteration independent constants. If additionally the sequence of error terms* $\{\varepsilon_i\}_{i \geq 1}$ *is square-summable, we have that* $\mathsf{G}(\boldsymbol{x}^t) \xrightarrow{\text{a.s.}} \mathbf{0}$ *as* $t \to \infty$.

Proof. To simplify our notations and analysis we will use $\gamma = 1/(\alpha L_{\mathsf{max}})$ with $\alpha > 1$. Note that Assumption 3 implies that there exists $f^* > -\infty$ such that we have almost surely $f^* \leq f(\boldsymbol{x}^k)$, $k \geq 1$. Consider the iteration k of BC-PnP in (6), where the random variables i_k are selected uniformly at random from $\{1, \cdots, b\}$. This implies that

$$\mathbb{E}\left[\|\boldsymbol{x}^k - \boldsymbol{x}^{k-1}\|_2^2 \,|\, \boldsymbol{x}^{k-1}\right] = \frac{1}{b}\sum_{j=1}^{b} \|\boldsymbol{x}_j^{k-1} - \mathsf{D}_{\sigma_j}(\boldsymbol{x}_j^{k-1} - \gamma \nabla_j g(\boldsymbol{x}^{k-1}))\|_2^2 \tag{17}$$

$$= \frac{\gamma^2}{b}\sum_{j=1}^{b} \|\mathsf{G}_j(\boldsymbol{x}^{k-1})\|_2^2 = \frac{\gamma^2}{b}\|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2 = \frac{1}{b(\alpha L_{\mathsf{max}})^2}\|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2.$$

On the other hand, from Lemma 2, we have almost surely that

$$f(\boldsymbol{x}^k) \leq f(\boldsymbol{x}^{k-1}) - (\alpha - 1)\frac{L_{\mathsf{max}}}{2}\|\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}\|_2^2 + \frac{\lambda \varepsilon_k^2}{2}$$

By taking conditional expectation of this bound, subtracting f^* from both sides, and using the equality (17), we get

$$\mathbb{E}\left[\left(f(\boldsymbol{x}^k) - f^*\right) \,|\, \boldsymbol{x}^{k-1}\right] \leq \left(f(\boldsymbol{x}^{k-1}) - f^*\right) - \theta\|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2 + \frac{\lambda\varepsilon_k^2}{2}, \tag{18}$$

where $\theta \coloneqq (\alpha - 1)/(2\alpha^2 b L_{\mathsf{max}})$. Hence, by averaging over $t \geq 1$ iterations and taking the total expectation, we obtain

$$\mathbb{E}\left[\frac{1}{t}\sum_{k=1}^{t}\|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2\right] \leq \frac{D_1}{t}(f(\boldsymbol{x}^0) - f^*) + D_2\left[\frac{1}{t}\sum_{k=1}^{t}\varepsilon_k^2\right],$$

where $D_1 \coloneqq 1/\theta$ and $D_2 \coloneqq \lambda/(2\theta)$. If $\{\varepsilon_k\}_{k\geq 1}$ in (18) is square summable, we can apply the supermartingale convergence theorem (see Section D), to get almost surely

$$\sum_{k=1}^{\infty}\|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2 < \infty,$$

which implies that $\|\mathsf{G}(\boldsymbol{x}^k)\|_2 \xrightarrow{\text{a.s.}} 0$ as $k \to \infty$.

□

C Useful technical lemmas

Lemma 1. *Consider the iteration $k \geq 1$ of BC-PnP under Assumptions 1-5 using the sequential block selection and the step-size $\gamma = 1/(\alpha L_{\mathsf{max}})$ with $\alpha > 1$. Then, we have that*

$$\|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2^2 \leq \frac{2}{(\alpha - 1)L_{\mathsf{max}}}\left(f(\boldsymbol{v}^{k-1}) - f(\boldsymbol{v}^k)\right) + \frac{\lambda}{2}\sum_{i=1}^{b}\varepsilon_{(k-1)b+i}^2, \quad k \geq 1,$$

where $\boldsymbol{v}^k \coloneqq (\boldsymbol{x}_1^{(k-1)b+1}, \cdots, \boldsymbol{x}_b^{kb})$, $f = g + h$ with h defined in (11)*, and $\lambda \coloneqq (\alpha L_{\mathsf{max}} + M_{\mathsf{max}})$.*

Proof. First note that due to the sequential nature of block updates, we have that

$$\boldsymbol{v}^k = \boldsymbol{x}^{kb} = (\boldsymbol{x}_1^{(k-1)b+1}, \cdots, \boldsymbol{x}_b^{kb}),$$
$$\boldsymbol{v}^{k-1} = \boldsymbol{x}^{(k-1)b} = \left(\boldsymbol{x}_1^{(k-2)b+1}, \cdots, \boldsymbol{x}_b^{(k-1)b}\right) = \left(\boldsymbol{x}_1^{(k-1)b}, \cdots, \boldsymbol{x}_b^{kb-1}\right).$$

By combining the observation above with Lemma 2, we get the following bound

$$(\alpha - 1)\frac{L_{\mathsf{max}}}{2}\sum_{i=1}^{b}\|\boldsymbol{x}_i^{(k-1)b+i} - \boldsymbol{x}_i^{(k-1)b+i-1}\|_2^2 = (\alpha - 1)\frac{L_{\mathsf{max}}}{2}\|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_2^2$$

$$\leq f(\boldsymbol{x}^{(k-1)b}) - f(\boldsymbol{x}^{kb}) + \frac{\lambda}{2}\sum_{i=1}^{b}\varepsilon_{(k-1)b+i}^2 = f(\boldsymbol{v}^{k-1}) - f(\boldsymbol{v}^k) + \frac{\lambda}{2}\sum_{i=1}^{b}\varepsilon_{(k-1)b+i}^2,$$

which directly leads to the desired result. □

Lemma 2. *Consider the iteration $k \geq 1$ of BC-PnP in* (6) *with the step-size $\gamma = 1/(\alpha L_{\mathsf{max}})$ with $\alpha > 1$. If Assumptions 1-5 are true, we have that*

$$f(\boldsymbol{x}^k) \leq f(\boldsymbol{x}^{k-1}) - (\alpha - 1)\frac{L_{\mathsf{max}}}{2}\|\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}\|_2^2 + \frac{\lambda\varepsilon_k^2}{2}, \tag{19}$$

where $f = g + h$, with h defined in (11)*, and $\lambda \coloneqq (\alpha L_{\mathsf{max}} + M_{\mathsf{max}})$.*

Proof. Consider the update that would be obtained by using the exact MMSE denoiser $\mathsf{D}_{\sigma_i}^*$

$$\boldsymbol{x}_j^* = \begin{cases} \boldsymbol{x}_j^{k-1} & \text{when } j \neq i_k \\ \mathsf{D}_{\sigma_j}^*(\boldsymbol{x}_j^{k-1} - \gamma\nabla_j g(\boldsymbol{x}^{k-1})) & \text{when } j = i_k \end{cases}, \quad j \in \{1, \cdots, b\},$$

From the fact that $\mathsf{D}_{\sigma_i}^*$ is the MMSE denoiser, we know that $\boldsymbol{x}_{i_k}^* \in \mathsf{Im}_\varepsilon(\mathsf{D}_{\sigma_{i_k}}^*)$ minimizes

$$\varphi_{i_k}(\boldsymbol{u}) \coloneqq \frac{1}{2\gamma}\|\boldsymbol{u} - (\boldsymbol{x}_{i_k}^{k-1} - \gamma\nabla_{i_k} g(\boldsymbol{x}^{k-1})\|_2^2 + h_{i_k}(\boldsymbol{u}), \quad \boldsymbol{u} \in \mathbb{R}^{n_i}. \tag{20}$$

From Assumption 5, we know that ∇h_{i_k} is M_{i_k}-Lipscthitz continuous over $\mathsf{Im}_\varepsilon(\mathsf{D}_{\sigma_{i_k}}^*)$, which implies

$$\|\nabla\varphi_{i_k}(\boldsymbol{u}) - \nabla\varphi_{i_k}(\boldsymbol{v})\|_2 \leq \lambda\|\boldsymbol{u} - \boldsymbol{v}\|_2 \quad \text{with} \quad \lambda \coloneqq (\alpha L_{\mathsf{max}} + M_{\mathsf{max}}),$$

for all $\boldsymbol{u}, \boldsymbol{v} \in \mathsf{Im}_\varepsilon(\mathsf{D}_{\sigma_{i_k}}^*)$. From the smoothness of φ_{i_k} and since $\boldsymbol{x}_{i_k}^*$ minimizes it, we have that

$$\varphi_{i_k}(\boldsymbol{x}_{i_k}^k) \leq \varphi_{i_k}(\boldsymbol{x}_{i_k}^*) + \nabla\varphi_{i_k}(\boldsymbol{x}_{i_k}^*)^{\mathsf{T}}(\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^*) + \frac{\lambda}{2}\|\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^*\|_2^2 \leq \varphi_{i_k}(\boldsymbol{x}_{i_k}^*) + \frac{\lambda\varepsilon_k^2}{2},$$

where in the second inequality we used the bound on the denoisers in Assumption 4. We thus have

$$\begin{aligned}
\varphi_{i_k}(\boldsymbol{x}_{i_k}^k) &= \frac{1}{2\gamma}\|\boldsymbol{x}_{i_k}^k - (\boldsymbol{x}_{i_k}^{k-1} - \gamma\nabla_{i_k} g(\boldsymbol{x}^{k-1}))\|_2^2 + h_{i_k}(\boldsymbol{x}_{i_k}^k) \\
&\leq \min_{\boldsymbol{u}\in\mathbb{R}^{n_i}} \left\{ \frac{1}{2\gamma}\|\boldsymbol{u} - (\boldsymbol{x}_{i_k} - \gamma\nabla_{i_k} g(\boldsymbol{x}_{i_k}^{k-1}))\|_2^2 + h_{i_k}(\boldsymbol{u}) \right\} + \frac{\lambda\varepsilon_k^2}{2} \\
&\leq \frac{1}{2\gamma}\|\boldsymbol{x}_{i_k}^{k-1} - (\boldsymbol{x}_{i_k}^{k-1} - \gamma\nabla_{i_k} g(\boldsymbol{x}^{k-1}))\|_2^2 + h_{i_k}(\boldsymbol{x}_{i_k}^{k-1}) + \frac{\lambda\varepsilon_k^2}{2}.
\end{aligned}$$

By expanding the first term on the left side of the inequality and simplifying, we obtain

$$h_{i_k}(\boldsymbol{x}_{i_k}^k) \leq h_i(\boldsymbol{x}_{i_k}^{k-1}) - \nabla_{i_k} g(\boldsymbol{x}^{k-1})^{\mathsf{T}}(\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}) - \frac{1}{2\gamma}\|\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}\|_2^2 + \frac{\lambda\varepsilon_k^2}{2}. \tag{21}$$

From the smoothness of g, we also have

$$g(\boldsymbol{x}^k) \leq g(\boldsymbol{x}^{k-1}) + \nabla_{i_k} g(\boldsymbol{x}^{k-1})^{\mathsf{T}}(\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}) + \frac{L_{\mathsf{max}}}{2}\|\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}\|_2^2. \tag{22}$$

By combining (21) and (22), and setting $\gamma = 1/(\alpha L_{\mathsf{max}})$, we get

$$\begin{aligned}
f(\boldsymbol{x}^k) &= g(\boldsymbol{x}^k) + h(\boldsymbol{x}^k) && (23)\\
&\leq g(\boldsymbol{x}^{k-1}) + \nabla_{i_k} g(\boldsymbol{x}^{k-1})^{\mathsf{T}}(\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}) + \frac{L_{\mathsf{max}}}{2}\|\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}\|_2^2 && (24)\\
&\quad + h(\boldsymbol{x}^{k-1}) - \nabla_{i_k} g(\boldsymbol{x}^{k-1})^{\mathsf{T}}(\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}) - \frac{1}{2\gamma}\|\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}\|_2^2 + \frac{\lambda\varepsilon_k^2}{2} && (25)\\
&= f(\boldsymbol{x}^{k-1}) - (\alpha - 1)\frac{L_{\mathsf{max}}}{2}\|\boldsymbol{x}_{i_k}^k - \boldsymbol{x}_{i_k}^{k-1}\|_2^2 + \frac{\lambda\varepsilon_k^2}{2}, && (26)
\end{aligned}$$

where we used the fact that $\boldsymbol{x}_j^k = \boldsymbol{x}_j^{k-1}$ for all $j \neq i_k$. □

Lemma 3. *Suppose Assumptions 1, 4, and 5 are true. We then have*

$$\|\nabla h(\boldsymbol{x}) - \nabla h(\boldsymbol{z})\|_2 \leq M_{\mathsf{max}}\|\boldsymbol{x} - \boldsymbol{z}\|_2, \quad \boldsymbol{x}, \boldsymbol{z} \in \mathsf{Im}_\varepsilon(\mathsf{D}_{\boldsymbol{\sigma}}^*). \tag{27}$$

Proof. Consider $\boldsymbol{x}, \boldsymbol{z} \in \mathsf{Im}_\varepsilon(\mathsf{D}_{\boldsymbol{\sigma}}^*)$. From the M_i-Lipschitz continuity of each ∇h_i, we get

$$\|\nabla h(\boldsymbol{x}) - \nabla h(\boldsymbol{z})\|_2^2 = \sum_{i=1}^b \|\nabla h_i(\boldsymbol{x}_i) - \nabla h_i(\boldsymbol{z}_i)\|_2^2 \leq \sum_{i=1}^b M_i^2\|\boldsymbol{x}_i - \boldsymbol{z}_i\|_2^2 \leq {M_{\mathsf{max}}}^2\|\boldsymbol{x} - \boldsymbol{z}\|_2^2.$$

□

D Background material

D.1 Supermartingale convergence theorem

Our analysis of the randomized BC-PnP algorithm relies on the classical result from the probability theory known as Supermargingale Convergence Theorem. The theorem is extensively used in the optimization literature (see Appendix A in [86] and Proposition 2 in [83]).

Theorem (Supermartingale theorem). *Let F^k, G^k, and E^k, be three sequences of random variables and let $\mathcal{F}_k$ be sets of random variables such that $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$ for all $k \geq 1$. Assume that*

- *F^k, G^k, and E^k are functions of the random variables in $\mathcal{F}_k$. Additionally, $F^k \geq 0$, $G^k \geq 0$, and $E^k \geq 0$ almost surely for $k \geq 1$.*
- *For each $k \geq 1$, we have*

$$\mathbb{E}[F^k \,|\, \mathcal{F}_{k-1}] \leq F^{k-1} - G^{k-1} + E^{k-1}.$$

- *We have almost surely*

$$\sum_{k=0}^{\infty} E^k < \infty.$$

Then, we have almost surely

- *$\sum_{k=1}^{\infty} G^{k-1} < \infty$;*
- *$F^k \to F^{\infty}$, where F^{∞} is a nonegative random variable.*

D.2 MMSE denoising as proximal operator

The relationship between MMSE estimation and regularized inversion has been established by Gribonval in [66] and has been discussed in other contexts [82, 87, 88]. This relationship was formally connected to PnP methods in [23], leading to their new interpretation for MMSE denoisers. In this section, we review the key argument connecting MMSE denoising and proximal operators.

It is well known that the estimator (8) can be compactly expressed using the *Tweedie's formula*

$$\mathsf{D}^*_{\sigma_i}(\boldsymbol{z}_i) = \boldsymbol{z}_i - \sigma_i^2 \nabla h_{\sigma_i}(\boldsymbol{z}_i) \quad \text{with} \quad h_{\sigma_i}(\boldsymbol{z}_i) = -\log(p_{\boldsymbol{z}_i}(\boldsymbol{z}_i)), \tag{28}$$

which can be obtained by differentiating (8) using the expression for the probability distribution

$$p_{\boldsymbol{z}_i}(\boldsymbol{z}_i) = (p_{\boldsymbol{x}_i} * \phi_{\sigma_i})(\boldsymbol{z}_i) = \int_{\mathbb{R}^{n_i}} \phi_{\sigma_i}(\boldsymbol{z}_i - \boldsymbol{x}_i) p_{\boldsymbol{x}_i}(\boldsymbol{x}_i)\, \mathsf{d}\boldsymbol{x}_i, \tag{29}$$

where

$$\phi_{\sigma_i}(\boldsymbol{x}_i) \coloneqq \frac{1}{(2\pi\sigma_i^2)^{\frac{n_i}{2}}} \exp\left(-\frac{\|\boldsymbol{x}_i\|^2}{2\sigma_i^2}\right).$$

Since ϕ_{σ_i} is infinitely differentiable, so are $p_{\boldsymbol{z}_i}$ and $\mathsf{D}^*_{\sigma_i}$. By differentiating $\mathsf{D}^*_{\sigma_i}$, one can show that the Jacobian of $\mathsf{D}^*_{\sigma_i}$ is positive definite (see Lemma 2 in [66])

$$\mathsf{J}\mathsf{D}^*_{\sigma_i}(\boldsymbol{z}_i) = \mathbf{I} - {\sigma_i}^2 \mathsf{H} h_{\sigma_i}(\boldsymbol{z}_i) \succ 0, \quad \boldsymbol{z}_i \in \mathbb{R}^{n_i}, \tag{30}$$

where $\mathsf{H}h_{\sigma_i}$ denotes the Hessian matrix of the function h_{σ_i}. Finally, Assumption 1 also implies that $\mathsf{D}^*_{\sigma_i}$ is a *one-to-one* mapping from $\mathbb{R}^{n_i}$ to $\mathsf{Im}(\mathsf{D}^*_{\sigma_i})$, which means that $(\mathsf{D}^*_{\sigma_i})^{-1} : \mathsf{Im}(\mathsf{D}^*_{\sigma_i}) \to \mathbb{R}^{n_i}$ is well defined and also infinitely differentiable over $\mathsf{Im}(\mathsf{D}^*_{\sigma_i})$ (see Lemma 1 in [66]). This directly implies that the regularizer h_i in (12) is also infinitely differentiable for any $\boldsymbol{x}_i \in \mathsf{Im}(\mathsf{D}^*_{\sigma_i})$.

We will now show that

$$\mathsf{D}^*_{\sigma_i}(\boldsymbol{z}_i) = \mathsf{prox}_{\gamma h_i}(\boldsymbol{z}_i) = \operatorname*{\mathsf{arg\,min}}_{\boldsymbol{x} \in \mathbb{R}^{n_i}} \left\{ \frac{1}{2}\|\boldsymbol{x}_i - \boldsymbol{z}_i\|^2 + \gamma h_i(\boldsymbol{x}_i) \right\}$$

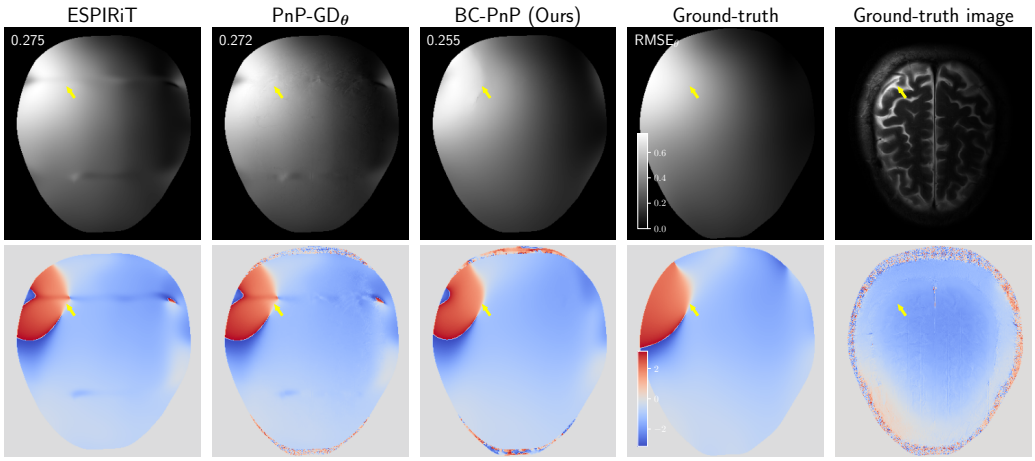

Figure 4: *Illustration of estimated CSM from several methods on CS-PMRI with the sampling factor $R = 6$. The top and the bottom rows are the magnitude and the phase of the CSMs, respectively. The quantities in the top-left corner of each image in the top row provide RMSE values for each method. Ground-truth image was obtained using the fully sampled data corresponding to the ground truth CSMs. This figure highlights the effectiveness of BC-PnP for estimating the measurement operator.*

where h_i is a (possibly nonconvex) function defined in (12). Our aim is to show that $\boldsymbol{u}^* = \boldsymbol{z}_i$ is the unique stationary point and global minimizer of

$$\varphi(\boldsymbol{u}) \coloneqq \frac{1}{2}\|\mathsf{D}^*_{\sigma_i}(\boldsymbol{u}) - \boldsymbol{z}_i\|^2 + \gamma h_i(\mathsf{D}^*_{\sigma_i}(\boldsymbol{u})), \quad \boldsymbol{u} \in \mathbb{R}^{n_i}.$$

By using the definition of h_i in (12) and the Tweedie's formula (28), we get

$$\varphi(\boldsymbol{u}) = \frac{1}{2}\|\mathsf{D}^*_{\sigma_i}(\boldsymbol{u}) - \boldsymbol{z}_i\|^2 - \frac{\sigma_i^4}{2}\|\nabla h_{\sigma_i}(\boldsymbol{u})\|^2 + \sigma_i^2 h_{\sigma_i}(\boldsymbol{u}).$$

The gradient of φ is then given by

$$\nabla\varphi(\boldsymbol{u}) = [\mathsf{JD}^*_{\sigma_i}(\boldsymbol{u})](\mathsf{D}^*_{\sigma_i}(\boldsymbol{u}) - \boldsymbol{z}_i) + \sigma_i^2[\mathbf{I} - \sigma_i^2 \mathsf{H} h_{\sigma_i}(\boldsymbol{u})]\nabla h_{\sigma_i}(\boldsymbol{u}) = [\mathsf{JD}^*_{\sigma_i}(\boldsymbol{u})](\boldsymbol{u} - \boldsymbol{z}_i),$$

where we used (30) in the second line and (28) in the third line. Now consider a scalar function $q(\nu) = \varphi(\boldsymbol{z}_i + \nu\boldsymbol{u})$ and its derivative

$$q'(\nu) = \nabla\varphi(\boldsymbol{z}_i + \nu\boldsymbol{u})^{\mathsf{T}}\boldsymbol{u} = \nu\boldsymbol{u}^{\mathsf{T}}[\mathsf{JD}^*_{\sigma_i}(\boldsymbol{z}_i + \nu\boldsymbol{u})]\boldsymbol{u}.$$

Positive definiteness of the Jacobian (30) implies $q'(\nu) < 0$ and $q'(\nu) > 0$ for $\nu < 0$ and $\nu > 0$, respectively. Thus, $\nu = 0$ is the global minimizer of q. Since $\boldsymbol{u} \in \mathbb{R}^{n_i}$ is an arbitrary vector, we have that φ has no stationary point beyond $\boldsymbol{u}^* = \boldsymbol{z}_i$ and that $\varphi(\boldsymbol{z}_i) < \varphi(\boldsymbol{u})$ for any $\boldsymbol{u} \neq \boldsymbol{z}_i$.

E Additional Technical Details

We present some technical details and results that were omitted from the main paper. We used the following root mean squared error (RMSE) for quantitatively comparing different algorithms

$$\mathsf{RMSE}(\widehat{\boldsymbol{z}}, \boldsymbol{z}) = \frac{\|\widehat{\boldsymbol{z}} - \boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2} \tag{31}$$

where $\widehat{\boldsymbol{z}}$ and $\boldsymbol{z}$ represents the estimation and ground truth respectively. We ran BC-PnP and its ablated variants using a maximum number of 500 iterations with the stopping criterion measuring the relative norm difference between iterations to be less than 10^{-5}. We trained denoisers for images and measurement operators to optimize the MSE loss by using the Adam [89] optimizer. We set the learning rate of Adam to 10^{-5}. We conducted all experiments on a machine equipped with an AMD Ryzen Threadripper 3960X 24-Core Processor and 4 NVIDIA GeForce RTX 3090 GPUs.

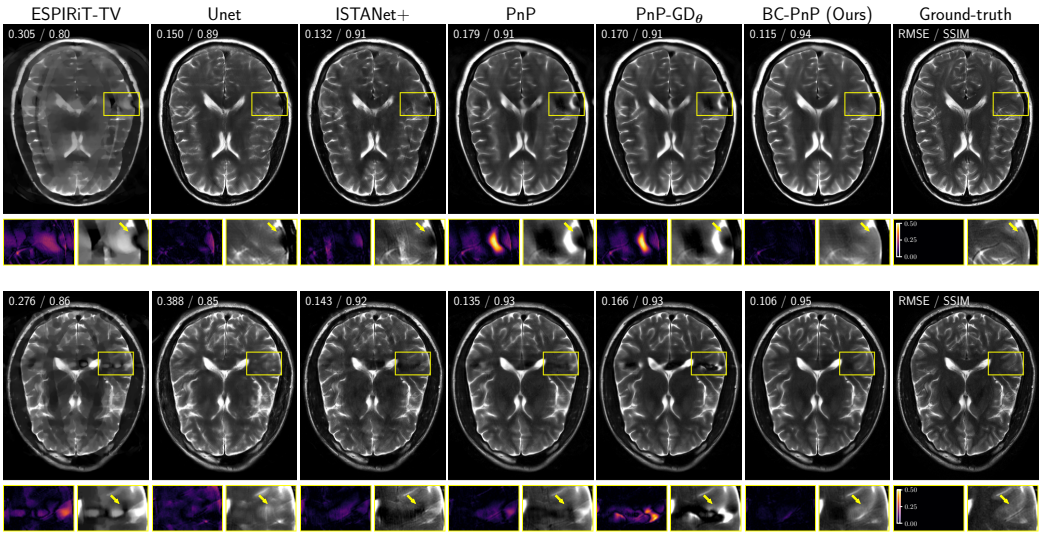

Figure 5: *Illustration of results from several well-known methods, including those were omitted from the main paper, on CS-PMRI with the sampling factor $R = 8$ (**top row**) and $R = 6$ (**bottom row**). The quantities in the top-left corner of each image provide the RMSE and SSIM values for each method. The squares at the bottom of each image shows the error and the corresponding zoomed area in the image. Note the excellent performance of BC-PnP that uses a learned deep denoiser on the CSMs.*

E.1 Additional Details for CS-PMRI

Figure 4 illustrates the visual results of the estimated CSM for an acceleration factor of $R = 6$. The widely used ESPIRiT algorithm estimates CSM directly from the ACS of the raw measurement, leading to imaging artifacts highlighted by yellow arrows under a high acceleration factor. Although PnP-GD$_{\boldsymbol{\theta}}$ can reduce such imaging artifacts by jointly estimating images and CSMs, its performance is suboptimal compared to BC-PnP. Figure 4 shows the effectiveness and superior performance of BC-PnP in estimating CSMs, which we attribute to its ability to use a DL denoiser as the CSM prior.

Figure 5 visually illustrates results from several well-known methods, including those were omitted from the main paper, on CS-PMRI with acceleration factors $R = 8$ and $R = 6$. ESPIRiT-TV leads to the loss of details due to the well-known "staircasing effect". While Unet can outperform ESPIRiT-TV by learning a prior end-to-end from a training dataset, its performance is suboptimal compared with ISTA-Net+ that incorporates the pre-estimated measurement operator into the network architecture. PnP and PnP-GD$_{\boldsymbol{\theta}}$ use pre-trained DL denoiser as image priors, leading competitive performance against ISTA-Net+. Figure 5 demonstrates that BC-PnP can achieve quantitatively and qualitatively superior performance over several baselines by jointly estimating images and CSMs.

Figure 6 shows ground truth MR images corresponding to the fully-sampled data that was used to generate measurement on the CS-PMRI experiments.

E.2 Additional Details for Blind Image Deblurring

Figure 7 presents visual results from several well-known methods, including those were omitted from the main paper, on blind image deblurring with the Gaussian kernel. Pan-DCP estimates a deblur kernel from the blurry measurement and then reconstructs the image using a non-DL image prior. SelfDeblur jointly trains two deep image priors (DIPs) on the image and the blur kernel, respectively, but note how its reconstructions are translated compared to the ground truth. DeblurGANv2 enables debluring of an image without the knowledge of the blur kernel, but its performance is noticeably suboptimal. While USRNet reconstructs sharp images given a pre-estimated kernel, the details in the corresponding images are inconsistent relative to the ground truth (see the texture of the tiger skin highlighted by yellow arrows). Note how BC-PnP using a deep denoiser on the unknown kernel outperforms several baselines and matches the performance of PnP that knows the true kernel.

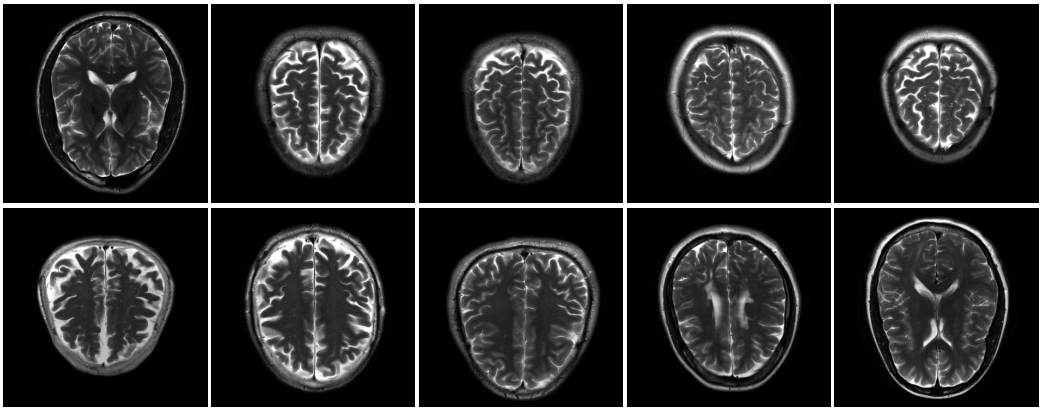

Figure 6: *Ground truth images that were used to generate the measurements in CS-PMRI.*

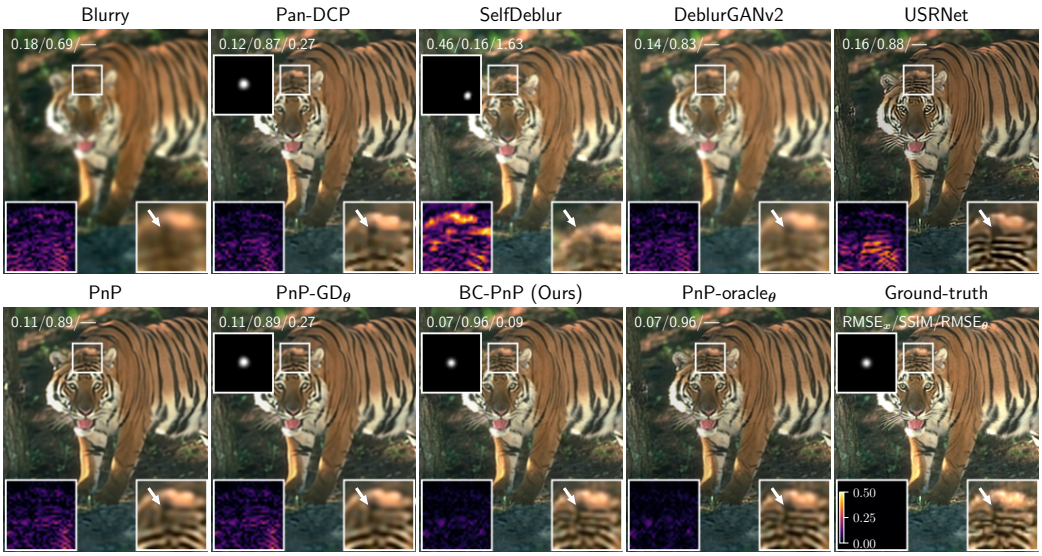

Figure 7: *Results from several well-known methods, including those were omitted from the main paper, on blind image deblurring with the Gaussian kernel. The squares at the top of each image show the estimated kernels. The quantities in the top-left corner of each image provide the RMSE and SSIM values for each method. The squares at the bottom of each image highlight the error and the corresponding zoomed image region. Note how BC-PnP using a deep denoiser on the unknown kernel performs as well as the oracle PnP that knows the ground truth kernel. Note also the effectiveness of BC-PnP for estimating the blur kernel.*

Figure 8 shows the images that were used to generate measurements for blind image deblurring.

Figure 8: *Ground truth images used for generating measurements for blind image deblurring.*