Supplementary Material for "Plug-and-Play Algorithms for Large-scale Snapshot Compressive Imaging"

Xin Yuan
Bell Labs
NJ USA
xyuan@bell-labs.com

Yang Liu
MIT
MA USA
yliu12@mit.edu

Jinli Suo, Qionghai Dai
Dept. of Automation & Institute for Brain and Cognitive Sciences, Tsinghua Univ., Beijing China
{jlsuo,daiqh}@tsinghua.edu.cn

We strongly encourage the readers to watch the supplementary videos of the reconstructed large-scale video data.

1. Proof of Lemma 1: Bounded Gradients of $f(\boldsymbol{x})$

Proof. The gradient of $f(\boldsymbol{x})$ in SCI is

$$\nabla f(\boldsymbol{x}) = \boldsymbol{H}^\top \boldsymbol{H}\boldsymbol{x} - \boldsymbol{H}^\top \boldsymbol{y}, \tag{1}$$

where $\boldsymbol{H}$ is a block diagonal matrix of size $n \times nB$.

- The $\boldsymbol{H}^\top \boldsymbol{y}$ is a non-negative constant since both the measurement $\boldsymbol{y}$ and the mask are non-negative in nature.
- Now let's focus on $\boldsymbol{H}^\top \boldsymbol{H}\boldsymbol{x}$. Since

$$\boldsymbol{H}^\top \boldsymbol{H} = \begin{bmatrix} \boldsymbol{D}_1 \\ \vdots \\ \boldsymbol{D}_B \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_1 \dots \boldsymbol{D}_B \end{bmatrix} \tag{2}$$

$$= \begin{bmatrix} \boldsymbol{D}_1^2 & \boldsymbol{D}_1\boldsymbol{D}_2 & \cdots & \boldsymbol{D}_1\boldsymbol{D}_B \\ \boldsymbol{D}_1\boldsymbol{D}_2 & \boldsymbol{D}_2^2 & \cdots & \boldsymbol{D}_2\boldsymbol{D}_B \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{D}_1\boldsymbol{D}_B & \boldsymbol{D}_2\boldsymbol{D}_B & \cdots & \boldsymbol{D}_B^2 \end{bmatrix}, \tag{3}$$

due to this special structure, $\boldsymbol{H}^\top \boldsymbol{H}\boldsymbol{x}$ is the weighted sum of the $\boldsymbol{x}$ and $\|\boldsymbol{H}^\top \boldsymbol{H}\boldsymbol{x}\|_2 \leq BC_{\max}\|\boldsymbol{x}\|_2$, where $C_{\max}$ is the maximum value in the sensing matrix. Usually, the sensing matrix is normalized to $[0, 1]$ and this leads to $C_{\max} = 1$ and therefore $\|\boldsymbol{H}^\top \boldsymbol{H}\boldsymbol{x}\|_2 \leq B\|\boldsymbol{x}\|_2$.

Thus, $\nabla f(\boldsymbol{x})$ is bounded. Furthermore,

- If the mask element $D_{i,j}$ is drawn from a binary distribution with entries {0,1} with a property of $p_1 \in (0, 1)$ being 1, then

$$\|\boldsymbol{H}^\top \boldsymbol{H}\boldsymbol{x}\|_2 \leq p_1 B\|\boldsymbol{x}\|_2 \tag{4}$$

with a high probability; usually, $p_1 = 0.5$ and thus $\|\boldsymbol{H}^\top \boldsymbol{H}\boldsymbol{x}\|_2 \leq 0.5B\|\boldsymbol{x}\|_2$.

- If the mask element $D_{i,j}$ is drawn from a Gaussian distribution $\mathcal{N}(0, \sigma^2)$ as in [2, 3], though it is not practical to get negative modulation (values of $D_{i,j}$) in hardware,

$$\|\boldsymbol{H}^\top \boldsymbol{H}\boldsymbol{x}\|_2 \leq B\sigma^2\|\boldsymbol{x}\|_2 \overset{\sigma=1}{=} B\|\boldsymbol{x}\|_2, \tag{5}$$

with a high probability.

□

2. Proof of Theorem 1: Fixed-point Convergence of Plug-and-Play GAP

Before we show the full proof, we first present 2 lemmas, which are important and based on the special structure of SCI sensing matrix.

Recall Assumption 1 in the main paper.

Assumption 1. *We assume that $\{R_j\}_{j=1}^n > 0$. This means for each spatial location j, the B frame modulation masks at this location have at least one non-zero entries. We further assume $R_{\max} > R_{\min}$.*

Lemma 1. *For any $\boldsymbol{x} \in \mathbb{R}^{nB}$, consider $D_{i,j}, i = 1, \dots, B; j = 1, \dots, n$ satisfy Assumption 1, then*

$$\begin{aligned} \frac{R_{\min}}{R_{\max}}\|\boldsymbol{x}\|_2^2 &\leq \|\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x}\|_2^2 \\ &= \|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{x}\|_2^2. \end{aligned} \tag{6}$$

Proof.

$$\|\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x}\|_2^2$$
$$= \boldsymbol{x}^\top \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x} \tag{7}$$
$$= \boldsymbol{x}^\top \boldsymbol{H}^\top \boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x} \tag{8}$$
$$= \|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{x}\|_2^2. \tag{9}$$

Further

$$\|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{x}\|_2^2 = \sum_{j=1}^{n}\left[\frac{\sum_{i=1}^{B} D_{i,j}x_{i,j}}{\sqrt{R_j}}\right]^2 \tag{10}$$

$$= \sum_{j=1}^{n}\frac{(D_{1,j}x_{1,j}+\cdots+D_{B,j}x_{B,j})^2}{R_j} \tag{11}$$

Recall that $R_j = \sum_{i=1}^{B} D_{i,j}^2 > 0$, and $D_{i,j} \geq 0$, $\forall i = 1,\ldots,B$ by Assumption 1. Furthermore, $\forall j$, at least one element in $\left\{\frac{D_{1,j}}{\sqrt{R_j}},\ldots,\frac{D_{B,j}}{\sqrt{R_j}}\right\}$ is larger than 0 and we denotes it by $\frac{D_{i^*,j}}{\sqrt{R_j}} > 0$. Following this,

$$\|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{x}\|_2^2 = \sum_{j=1}^{n}\frac{(D_{i,j}x_{i,j}+\cdots+D_{B,j}x_{B,j})^2}{D_{1,j}^2+\cdots+D_{B,j}^2} \tag{12}$$

$$= \sum_{j=1}^{n}\frac{(\sum_{i=1}^{B} D_{i,j}x_{i,j})^2}{\sum_{i=1}^{B} D_{i,j}^2} \leq \sum_{j=1}^{n}\sum_{i=1}^{B}\frac{D_{i,j}^2 x_{i,j}^2}{D_{i,j}^2} \tag{13}$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{B} x_{i,j}^2 = \|\boldsymbol{x}\|_2^2, \tag{14}$$

where (13) is from the Cauchy–Schwarz inequality by defining the following two vectors

$$\boldsymbol{d}_j = [D_{1,j},\ldots,D_{B,j}]^\top, \tag{15}$$

$$\boldsymbol{x}_j = [x_{1,j},\ldots,x_{B,j}]^\top, \tag{16}$$

and $\langle \boldsymbol{d}_j, \boldsymbol{x}_j\rangle^2 \leq \|\boldsymbol{d}_j\|_2^2\|\boldsymbol{x}_j\|_2^2$.

Since we have assumed $D_{i,j} \geq 0$ and at least one element in $\boldsymbol{d}_j$ is larger than 0, the equality $\|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2$ only happens when $\boldsymbol{x} = \boldsymbol{0}$ or $\boldsymbol{x}_j = c\boldsymbol{d}_j$, where c is a constant and this means $\boldsymbol{x}_j$ and $\boldsymbol{d}_j$ are linearly dependent.

This will be an extreme case and this means the signal is linearly dependent on the mask and thus the SCI system cannot resolve the signal.

Recall that $\boldsymbol{R} = \boldsymbol{H}\boldsymbol{H}^\top$ is a diagonal matrix and $\lambda_{\min}(\boldsymbol{H}\boldsymbol{H}^\top) = R_{\min}$, $\lambda_{\max}(\boldsymbol{H}\boldsymbol{H}^\top) = R_{\max}$. We have

$$R_{\min}\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{H}\boldsymbol{x}\|_2^2 \leq R_{\max}\|\boldsymbol{x}\|_2^2 \tag{17}$$

Following this and $\boldsymbol{R}^{-1} = \mathrm{diag}(R_1^{-1},\ldots,R_n^{-1})$, we have

$$\frac{R_{\min}}{R_{\max}}\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{x}\|_2^2 \leq \frac{R_{\max}}{R_{\min}}\|\boldsymbol{x}\|_2^2. \tag{18}$$

Along with (14), we have

$$\frac{R_{\min}}{R_{\max}}\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{x}\|_2^2, \tag{19}$$

which is the desired result. □

Lemma 2. *For any $\boldsymbol{x} \in \mathbb{R}^{nB}$, consider $D_{i,j}, i = 1,\ldots,B; j = 1,\ldots,n$ satisfy Assumption 1, then*

$$\|(\boldsymbol{I}-\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H})\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2 - \|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{x}\|_2^2. \tag{20}$$

Proof.

$$\|(\boldsymbol{I}-\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x}\|_2^2$$

$$= \|\boldsymbol{x}-\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x}\|_2^2 \tag{21}$$

$$= [\boldsymbol{x}-\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x}]^\top[\boldsymbol{x}-\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x}] \tag{22}$$

$$= \boldsymbol{x}^\top\boldsymbol{x} - 2\boldsymbol{x}^\top\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^\top\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x} \tag{23}$$

$$= \boldsymbol{x}^\top\boldsymbol{x} - \boldsymbol{x}^\top\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x} \tag{24}$$

$$= \|\boldsymbol{x}\|_2^2 - \|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{x}\|_2^2, \tag{25}$$

where we have used $\boldsymbol{R} = \boldsymbol{H}\boldsymbol{H}^\top$.

Along with Lemma 1, we have

$$\|(\boldsymbol{I}-\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}\boldsymbol{x}\|_2^2 \leq \left(1-\frac{R_{\min}}{R_{\max}}\right)\|\boldsymbol{x}\|_2^2. \tag{26}$$

□

Now we prove Theorem 1 in the main paper.

Proof. From (21) in the main paper,

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{v}^{(k)} + \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{v}^{(k)}), \tag{27}$$

we have

$$\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} = \boldsymbol{v}^{(k)} - \boldsymbol{x}^{(k)} + \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{v}^{(k)}), \tag{28}$$

Following this,

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|_2^2 \tag{29}$$

$$= \|\boldsymbol{v}^{(k)} + \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{v}^{(k)}) - \boldsymbol{x}^{(k)}\|_2^2 \tag{30}$$

$$= \|\boldsymbol{v}^{(k)} + \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}(\boldsymbol{H}\boldsymbol{x}^{(k)} - \boldsymbol{H}\boldsymbol{v}^{(k)}) - \boldsymbol{x}^{(k)}\|_2^2$$

$$= \|(\boldsymbol{I}-\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H})(\boldsymbol{v}^{(k)} - \boldsymbol{x}^{(k)})\|_2 \tag{31}$$

$$= \|\boldsymbol{v}^{(k)} - \boldsymbol{x}^{(k)}\|_2^2 - \|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}(\boldsymbol{v}^{(k)} - \boldsymbol{x}^{(k)})\|_2^2 \tag{32}$$

$$\leq \left(1-\frac{R_{\min}}{R_{\max}}\right)\|\boldsymbol{v}^{(k)} - \boldsymbol{x}^{(k)}\|_2^2, \tag{33}$$

where the last equation (33) comes from Lemma 2. Now let's use $\boldsymbol{v}^{(k)} = \mathcal{D}_{\sigma_{k-1}}(\boldsymbol{x}^{(k)})$ and the definition of bounded denoiser

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|_2^2 \leq \left(1-\frac{R_{\min}}{R_{\max}}\right)\|\boldsymbol{v}^{(k)} - \boldsymbol{x}^{(k)}\|_2^2 \tag{34}$$

$$= \left(1-\frac{R_{\min}}{R_{\max}}\right)\|\mathcal{D}_{\sigma_{k-1}}(\boldsymbol{x}^{(k)}) - \boldsymbol{x}^{(k)}\|_2^2 \tag{35}$$

$$\leq nBC\left(1-\frac{R_{\min}}{R_{\max}}\right)\sigma_{k-1}^2 \tag{36}$$

$$= nBC\left(1-\frac{R_{\min}}{R_{\max}}\right)\lambda_{k-1}, \tag{37}$$

where we have used $\sigma_{k-1} = \sqrt{\lambda_{k-1}}$.

Based on this,

$$\begin{aligned}
&\|\boldsymbol{v}^{(k+1)} - \boldsymbol{v}^{(k)}\|_2^2 \\
&= \|\mathcal{D}_{\sigma_k}(\boldsymbol{x}^{(k+1)}) - \mathcal{D}_{\sigma_{k-1}}(\boldsymbol{x}^{(k)})\|_2^2 && (38)\\
&= \|\mathcal{D}_{\sigma_k}(\boldsymbol{x}^{(k+1)}) - \boldsymbol{x}^{(k+1)} \\
&\quad + \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} + \boldsymbol{x}^{(k)} - \mathcal{D}_{\sigma_{k-1}}(\boldsymbol{x}^{(k)})\|_2^2 && (39)\\
&\leq 2\|\mathcal{D}_{\sigma_k}(\boldsymbol{x}^{(k+1)}) - \boldsymbol{x}^{(k+1)}\|_2^2 \\
&\quad + 2\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|_2^2 + 2\|\boldsymbol{x}^{(k)} - \mathcal{D}_{\sigma_{k-1}}(\boldsymbol{x}^{(k)})\|_2^2 \\
&\leq 2nBC\lambda_k + 2nBC\left(1 - \frac{R_{\min}}{R_{\max}}\right)\lambda_{k-1} + 2nBC\lambda_{k-1} && (40)\\
&\overset{\lambda_k \leq \lambda_{k-1}}{\leq} \left(6nBC - \frac{2nBCR_{\min}}{R_{\max}}\right)\lambda_{k-1}, && (41)
\end{aligned}$$

Following [1], we define $\boldsymbol{\theta}^{(k)} = (\boldsymbol{x}^{(k)}, \boldsymbol{v}^{(k)})$. Let Θ be the domain of $\boldsymbol{\theta}^{(k)}$ for all k. On $\boldsymbol{\Theta}$, we define a distance function $D : \boldsymbol{\Theta} \times \boldsymbol{\Theta} \to \mathbb{R}$ such that

$$D(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(j)}) = \frac{1}{nB}\left(\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(j)}\|_2^2 + \|\boldsymbol{v}^{(k)} - \boldsymbol{v}^{(j)}\|_2^2\right). \quad (42)$$

It follows from

$$\Delta_{k+1} = \tfrac{1}{\sqrt{nB}}\left(\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|_2 + \|\boldsymbol{v}^{(k+1)} - \boldsymbol{v}^{(k)}\|_2\right). \quad (43)$$

that $\Delta_{k+1} = D(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)})$. If we can show that $\{\boldsymbol{\theta}^{(k)}\}_{k=1}^{\infty}$ is a Cauchy sequence in $\boldsymbol{\Theta}$ with the distance function D, then $\boldsymbol{\theta}^{(k)}$ should converge.

Now we consider the following two cases of updating λ_k:

a) Monotone update rule by setting $\lambda_{k+1} = \xi\lambda_k$, with $\xi < 1$ for all k.

In this case, it is easily to see from Eq. (37) and Eq. (41) that

$$\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\|_2^2 \leq \left(7nBC - \frac{3nBCR_{\min}}{R_{\max}}\right)\lambda_0\xi^{k-1} \quad (44)$$

$$\overset{0<R_{\min}\leq R_{\max}}{\leq} 7nBC\lambda_0\xi^{k-1}. \quad (45)$$

As $\xi < 1$, when $k \to \infty$, $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\|_2 \to 0$. In other words, $\{\boldsymbol{\theta}^{(k)}\}_{k=1}^{\infty}$ is a Cauchy sequence and therefore, there must exists $\boldsymbol{\theta}^*$ such that

$$D(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^*) \to 0. \quad (46)$$

Consequently, we have $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_2^2 \to 0$ and $\|\boldsymbol{v}^{(k)} - \boldsymbol{v}^*\|_2^2 \to 0$.

This is fixed-point convergence.

b) Adaptive update via

i) If $\Delta_{k+1} \geq \eta\Delta_k$, then $\lambda_{k+1} = \xi\lambda_k$.

ii) If $\Delta_{k+1} < \eta\Delta_k$, then $\lambda_{k+1} = \lambda_k$.

In both cases, $\xi < 1$.

At k^{th} iteration, we have

- Case i) holds, from Eq. (37) and Eq. (41), we have
$$D(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)}) \leq 7C\lambda_{k-1}. \quad (47)$$
- Case ii) holds, since $\Delta_{k+1} = D(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)})$,
$$D(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)}) < \eta D(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(k-1)}). \quad (48)$$

As $k \to \infty$, one of the following situations will happen

S1: Case i) occurs infinite times but Case ii) occurs finite times.

S2: Case ii) occurs infinite times but Case i) occurs finite times.

S3: Both Case i) and Case ii) occur infinite times.

When S1 happens, there must exists a K_1 such that for $k \geq K_1$ only case i) will be visited. Following this,

$$D(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)}) \leq 7C\lambda_{K_1-1}\xi^{k-K_1-1}, \quad \forall k \geq K_1. \quad (49)$$

When S2 happens, there must exists a K_2 such that for $k \geq K_2$ only case ii) will be visited. Following this,

$$\begin{aligned}
D(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)}) &< \eta^{k-K_2} D(\boldsymbol{\theta}^{(K_2)}, \boldsymbol{\theta}^{(K_2-1)}) && (50)\\
&\leq 7C\lambda_{K_2-2}\eta^{k-K_2}, \ \forall k \geq K_2. && (51)
\end{aligned}$$

Combing (49) and (51), we have, for any $k \geq \max\{K_1, K_2\}$,

$$\begin{aligned}
&D(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)}) \\
&\leq 7C\max\left\{\lambda_{K_1-1}\xi^{-K_1-1}, \lambda_{K_2-2}\eta^{-K_2}\right\}(\max\{\xi, \eta\})^k && (52)
\end{aligned}$$

Let $\tilde{C} = 7C\max\left\{\lambda_{K_1-1}\xi^{-K_1-1}, \lambda_{K_2-2}\eta^{-K_2}\right\}$ and $\delta = \max\{\xi, \eta\}$, we have

$$D(\boldsymbol{\theta}^{(k+1)}, \boldsymbol{\theta}^{(k)}) \leq \tilde{C}\delta^k, \quad (53)$$

for some constant $\tilde{C}$ and $\delta < 1$.

Since S3 is a union of S1 and S2, we can verify that no matter which one of S1 to S3, for all integers $m > k$ and $k \to \infty$,

$$D(\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(k)}) \to 0. \quad (54)$$

Following this, $\{\boldsymbol{\theta}^{(k)}\}_{k=1}^{\infty}$ is a Cauchy sequence and thus it always converges. Therefore, there must exists $\boldsymbol{\theta}^*$ such that

$$D(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^*) \to 0. \quad (55)$$

Consequently, we have $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_2 \to 0$ and $\|\boldsymbol{v}^{(k)} - \boldsymbol{v}^*\|_2 \to 0$.

□

3. Proof of Theorem 2: Global Convergence of Plug-and-Play GAP

Proof. Again, let us start from (27), we have

$$\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^* = \boldsymbol{v}^{(k)} - \boldsymbol{x}^* + \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{v}^{(k)}) \tag{56}$$

$$= \boldsymbol{v}^{(k)} - \boldsymbol{x}^* + \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}(\boldsymbol{x}^* - \boldsymbol{v}^{(k)}) \tag{57}$$

Following the derivation in (32), we have

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\|_2^2 = \|\boldsymbol{v}^{(k)} - \boldsymbol{x}^*\|_2^2 - \|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{H}(\boldsymbol{v}^{(k)} - \boldsymbol{x}^*)\|_2^2 \tag{58}$$

$$\leq \left(1 - \frac{R_{\min}}{R_{\max}}\right)\|\boldsymbol{v}^{(k)} - \boldsymbol{x}^*\|_2^2. \tag{59}$$

This is

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\|_2$$

$$\leq \sqrt{1 - \frac{R_{\min}}{R_{\max}}}\|\boldsymbol{v}^{(k)} - \boldsymbol{x}^*\|_2 \tag{60}$$

$$\leq \sqrt{1 - \frac{R_{\min}}{R_{\max}}}\left(\|\boldsymbol{v}^{(k)} - \boldsymbol{x}^{(k)}\|_2 + \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_2\right) \tag{61}$$

$$\leq \sqrt{1 - \frac{R_{\min}}{R_{\max}}}\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_2 + \sqrt{nBC\sigma_{k-1}(1 - \frac{R_{\min}}{R_{\max}})} \tag{62}$$

$$= \sqrt{1 - \frac{R_{\min}}{R_{\max}}}\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_2 + \sqrt{nBC(1 - \frac{R_{\min}}{R_{\max}})}\lambda_{k-1}^{\frac{1}{4}} \tag{63}$$

$$= \sqrt{1 - \frac{R_{\min}}{R_{\max}}}\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_2 + \lambda_0\sqrt{nBC(1 - \frac{R_{\min}}{R_{\max}})}\xi^{\frac{k-1}{4}} \tag{64}$$

Note that since $\sqrt{1 - \frac{R_{\min}}{R_{\max}}} < 1$ and $\xi < 1$ and therefore, when $k \to \infty$,

$$\lambda_0\sqrt{nBC(1 - \frac{R_{\min}}{R_{\max}})}\xi^{\frac{k-1}{4}} \to 0. \tag{65}$$

In the main result, we have defined

$$C_g = \lambda_0\sqrt{nBC(1 - \frac{R_{\min}}{R_{\max}})}\xi^{-\frac{1}{4}}, \tag{66}$$

$$\varphi = \xi^{\frac{1}{4}}. \tag{67}$$

□

4. Simulation and Reconstruction Details

Benchmark data We follow the simulation setup and and use all the four benchmark datasets, *i.e.*, `Kobe`, `Traffic`, `Runner`, and `Drop` in [4]. We use two more datasets, *i.e.*, `Crash` and `Aerial` in [5][1] to cover a wider range of scenes. Full comparison of reconstruction frames of all seven PnP-GAP algorithms listed in Table 1 is shown in Fig. 1.

It can be seen clearly that PnP-FFDNet provides overall comparable results as the state-of-the-art (best among all the seven methods listed here) method DeSCI, as shown in Fig. 1 with significantly reduced running time (3 seconds vs. 103 minutes), as shown in Table 1 in the main paper.

Large-scale data Similar to the benchmark data, we simulate the color video SCI measurements for large-scale data with four YouTube slow-motion videos, *i.e.*, `Messi`[2], `Hummingbird`[3], `Swinger`[4], and `Football`[5].

The color video SCI system and sensing process follows the color video and depth SCI system in [8]. The scematic of a color video SCI system is shown in Fig. 2. A sequence of color scene is coded by the corresponding shifted random binary masks at each time step and finally summed up to form a shapshot measurement on the color Bayer RGB sensor (with a "RGGB" Bayer color filter array here).

For reconstruction, the snapshot measurement is splitted into four "RGGB" sub-measurements according to the Bayer pattern. These sub-measurements are reconstructed measurement-by-measurement following the gray-scale reconstruction process by iteratively update the signal in data domain (using GAP or ADMM) and prior domain (using plug-and-play denoisers). Finally, the reconstructed sub-video-frames representing different color channels (R, G1, G2, and B) are recombined to a mosaic image and then demosaiced to form full-color video frames. Note that for simulation of large-scale data using YouTube videos, we do not have the access to the raw video data before demosaicing, so we simply "up-sample" it by putting each color channel as the mosaic R, G1, G2, and B channels. In this way, there are two identical G channels here and the reconstructed and the size of demosaiced image is doubled (both in width and height). For example, for UHD color video `Football` with original image size of 3840×1644, the reconstructed video frames have the size of 7680×3288 (demosaiced). To make the readers less confusing, we simply call it UHD here

[1]The results of DeSCI (GAP-WNNM) is different from those reported in [5] because of parameter settings of DeSCI, specifically the input estimated noise levels for each iteration stage. We use exactly the same parameters as the DeSCI paper [4], which is publicly available at https://github.com/liuyang12/DeSCI.

[2]https://www.youtube.com/watch?v=sbPrevs6Pd4

[3]https://www.youtube.com/watch?v=RtUQ_pz5wlo

[4]https://www.youtube.com/watch?v=cfnbyX9G5Rk

[5]https://www.youtube.com/watch?v=EGAuWZYe2No

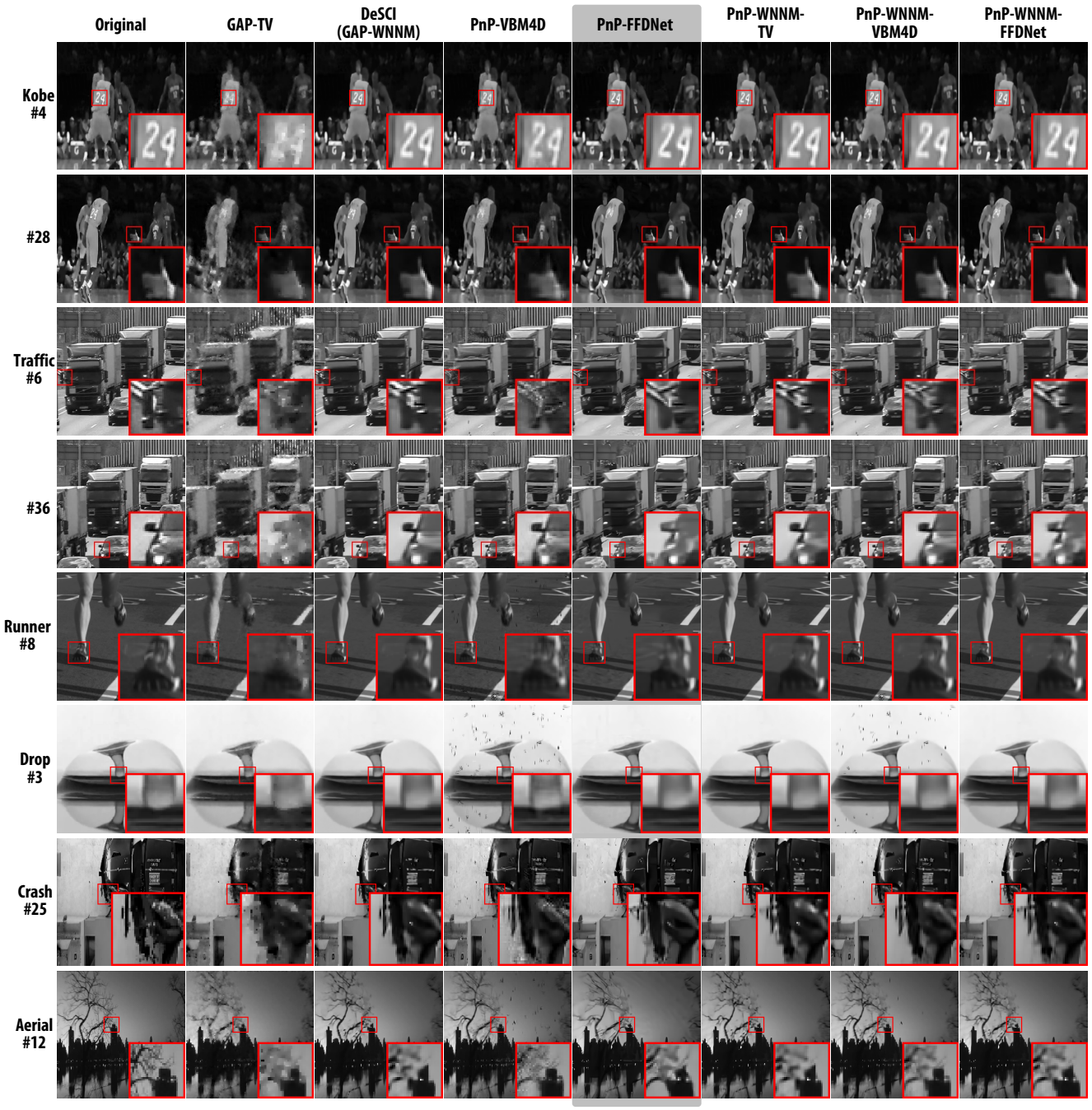

Figure 1. Full comparison of reconstructed frames of PnP-GAP algorithms (GAP-TV, DeSCI (GAP-WNNM), PnP-VBM4D, PnP-FFDNet, PnP-WNNM-TV, PnP-WNNM-VBM4D, and PnP-WNNM-FFDNet) on six simulated video SCI datasets.

(8K UHD exactly). And the quantitative metrics (PSNR and SSIM) are calculated before demosaicing.

We again strongly encourage the readers to watch the supplementary videos of the reconstructed large-scale video data.

- largescale_messi24.avi: A $1920 \times 1080 \times 24$ video reconstructed from a snapshot.
- largescale_hummingbird40.avi: A $1920 \times 1080 \times 40$ video reconstructed from a snapshot.
- largescale_swinger20.avi: A $3840 \times 2160 \times 20$ video reconstructed from a snapshot.
- largescale_football48.avi: A $3840 \times 1644 \times 48$ video reconstructed from a snapshot.

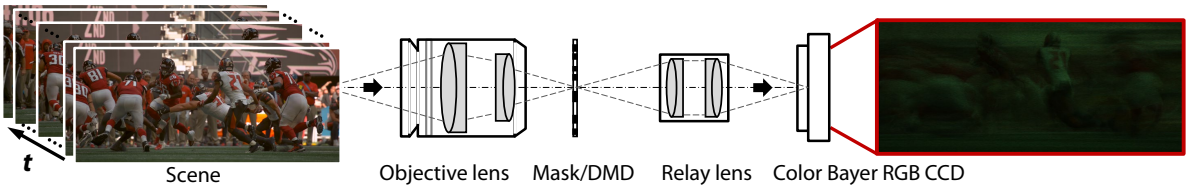

Figure 2. Schematic of a color video SCI system and its snapshot measurement (showing in Bayer RGB mode). A "RGGB" Bayer pattern is shown here.

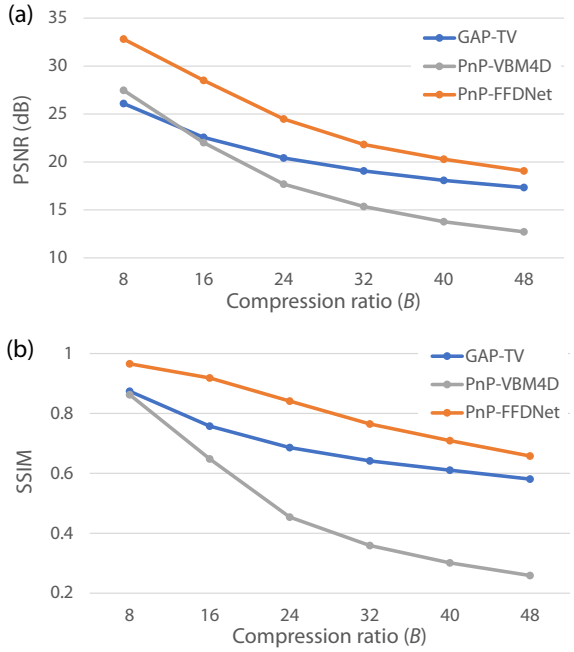

Figure 3. Reconstruction quality, *i.e.*, PSNR (a) and SSIM (b) varying compression ratios from 8 to 48. Higher is better.

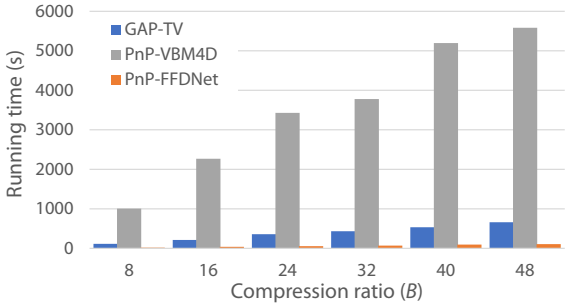

Figure 4. Running time (in seconds) varying compression ratios from 8 to 48. Lower is better.

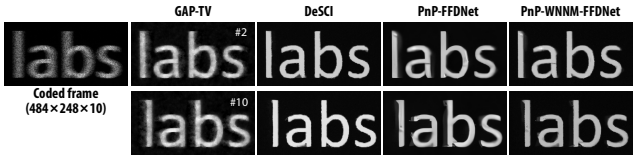

Figure 5. Real data: `labs` high-speed video SCI ($484 \times 248 \times 10$).

Performance varying compression ratios (B) In order to further illustrate the efficiency of the proposed PnP algorithms for SCI, especially in real SCI systems varying compression ratios (B), we show the reconstruction quality and speed of three PnP-SCI algorithms with compression ratios from 8 to 48 in Fig. 3 and Fig. 4, respectively. The data we used is the downsampled grayscale video of `Football` with pixel resoltion of 720p (1280×720). The other algorithms listed in Tab. 1 are too slow to be compared. And other deep-learning-based end-to-end networks, like [5, 6] would be not flexible to different compression ratios and require re-training the network for each compression ratio.

As we can see in Fig. 3 and Fig. 4, PnP-FFDNet is of the highest quality and fastest speed among these three fast PnP-SCI algorithms even with high compression ratios (up to 48). This further supports the idea that PnP-FFDNet would be the baseline for SCI reconstruction.

5. More Experimental Results

We show additional experimental results of the proposed PnP framework of real data captured by SCI cameras here. As mentioned in Tab. 2 of the main paper, we show the results of `labs` [6] and `UCF` [7] in Fig. 5 and Fig. 6, respectively.

Similar to the observation of the main paper, from Fig. 5 and Fig. 6, we can see that PnP-FFDNet, which only takes about 12 seconds for reconstruction, can provide comparable results as DeSCI, which needs hours even when performed in a frame-wise manner, as shown in Tab. 2. And PnP-FFDNet is significantly better than the speed runner-up GAP-TV in terms of motion-blur reduction and detail preservation, as shown in Figs. 5 and 6. Note that PnP-FFDNet is more than $5\times$ faster than GAP-TV in real datasets with regular size, and more than $7\times$ faster in large-scale datasets. In this way, PnP algorithms for SCI achieves a good balance of efficiency and flexibility and PnP-FFDNet could serve as a baseline for SCI recovery.

References

[1] Stanley H. Chan, Xiran Wang, and Omar A. Elgendy. Plug-and-play ADMM for image restoration: Fixed-point convergence

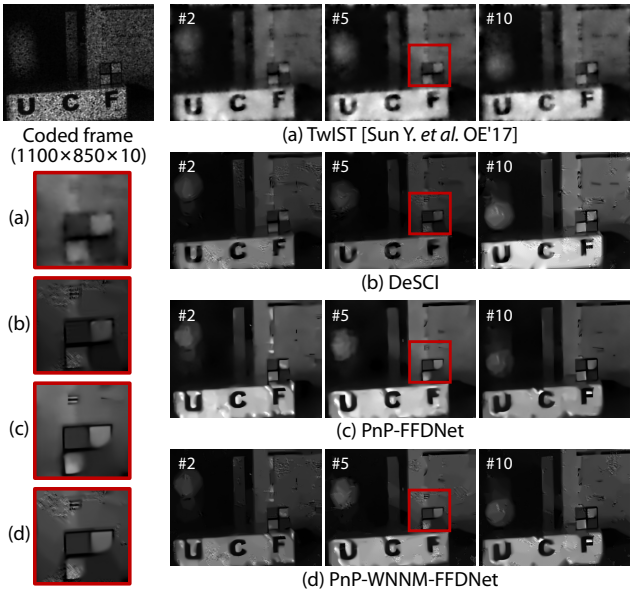

Figure 6. Real data: UCF high-speed video SCI ($1100 \times 850 \times 10$).

and applications. *IEEE Transactions on Computational Imaging*, 3:84–98, 2017.

[2] Shirin Jalali and Xin Yuan. Compressive imaging via one-shot measurements. In *IEEE International Symposium on Information Theory (ISIT)*, 2018.

[3] Shirin Jalali and Xin Yuan. Snapshot compressed sensing: Performance bounds and algorithms. *IEEE Transactions on Information Theory*, 65(12):8005–8024, Dec 2019.

[4] Yang Liu, Xin Yuan, Jinli Suo, David Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2990–3006, Dec 2019.

[5] Jiawei Ma, Xiaoyang Liu, Zheng Shou, and Xin Yuan. Deep tensor admm-net for snapshot compressive imaging. In *IEEE/CVF Conference on Computer Vision (ICCV)*, 2019.

[6] Mu Qiao, Ziyi Meng, Jiawei Ma, and Xin Yuan. Deep learning for video compressive sensing. *APL Photonics*, 5(3):030801, 2020.

[7] Yangyang Sun, Xin Yuan, and Shuo Pang. High-speed compressive range imaging based on active illumination. *Optics Express*, 24(20):22836–22846, Oct 2016.

[8] Xin Yuan, Patrick Llull, Xuejun Liao, Jianbo Yang, David J. Brady, Guillermo Sapiro, and Lawrence Carin. Low-cost compressive sensing for color video and depth. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3318–3325, 2014.