Multiple Sensor Image Fusion from Joint Adaptive Compressed Sampling

September 2, 2020

1 Introduction

Denote the hyperspectral image (HSI) and the paired multispectral image (MSI) to be $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times N_3}$ and $\mathcal{Z} \in \mathbb{R}^{N_1 \times N_2 \times n_3}$, respectively. The super-resolution image (SRI) is $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$. Let $\boldsymbol{X} \in \mathbb{R}^{(N_1 N_2) \times N_3}$, $\boldsymbol{Y} \in \mathbb{R}^{(n_1 n_2) \times N_3}$, $\boldsymbol{Z} \in \mathbb{R}^{(N_1 N_2) \times n_3}$ be the matricization of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. Suppose $\boldsymbol{Y} \approx \boldsymbol{P}_{12}\boldsymbol{X}$ for $\boldsymbol{P}_{12} \in \mathbb{R}^{(n_1 n_2) \times (N_1 N_2)}$ corresponding to the spatial blur. Also suppose $\boldsymbol{Z} \approx \boldsymbol{X}\boldsymbol{P}_3$ for $\boldsymbol{P}_3 \in \mathbb{R}^{N_3 \times n_3}$ corresponding to the spectral response of the multispectral sensor. The goal is to recover $\boldsymbol{X}$ from adaptive sampling of $\boldsymbol{Y}$ and $\boldsymbol{Z}$. In order to achieve this, we ask ourselves:

- $\mathcal{X}$, as a tensor, is sparse in what representation?
- How to design the sampling schemes for $\mathcal{Y}$ and $\mathcal{Z}$, such that the samples from the two sources contribute to the recovery of $\mathcal{X}$?

After being able to answer the above questions, there are further improvements we can think of, such as:

- Based on prior knowledge, only the discriminative bands needed to be recovered;
- Quantization of the measurements can be considered to make the work more practical.

A similar problem is considered in [6] with some success. The sparsifying system is the Kronecker product of 1D DCT and 2D DCT. The measurement matrices, taking values 0 and 1, are designed to be "balanced". A related problem is considered in [9]. For the data in fluorescent spectroscopy, it's found to be sparse in the Kronecker product of 1D Haar wavelet and 2D Haar wavelet. The measurements are subsamples of the original tensor multiplied by inverse Fourier transform in the spectral domain and 2D Hadamard transform in the spatial domain.

2 Problem Formulation

2.1 Compressed Sensing with Multiple Sensors

Suppose $\mathcal{X}$ has the following sparse representation, $vec(\boldsymbol{S}) = (\boldsymbol{\Phi}_3 \otimes \boldsymbol{\Phi}_{12}) \cdot vec(\boldsymbol{X})$. Put our problem in the context of compressed sensing with multiple sensors, we solve

$$\begin{aligned} &\min \ \|vec(\boldsymbol{S})\|_1 \\ &\text{s.t.} \begin{bmatrix} \boldsymbol{\Psi}_{12} \cdot \boldsymbol{I}_3 \otimes \boldsymbol{P}_{12} \\ \boldsymbol{\Psi}_3 \cdot \boldsymbol{P}_3^T \otimes \boldsymbol{I}_{12} \end{bmatrix} \cdot (\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T) \cdot vec(\boldsymbol{S}) \approx \begin{bmatrix} \boldsymbol{\Psi}_{12} \cdot vec(\boldsymbol{Y}) \\ \boldsymbol{\Psi}_3 \cdot vec(\boldsymbol{Z}) \end{bmatrix}. \end{aligned} \tag{1}$$

Denote $\boldsymbol{H}_{12} = \boldsymbol{I}_3 \otimes \boldsymbol{P}_{12}$ and $\boldsymbol{H}_3 = \boldsymbol{P}_3^T \otimes \boldsymbol{I}_{12}$ to be the sensor profiles. In the state of art,

- if $\boldsymbol{H}_{12}^T\boldsymbol{H}_{12} + \boldsymbol{H}_3^T\boldsymbol{H}_3 = \boldsymbol{I}$, guarantees about various measurement matrices are established in [3]. "Distributed sparsity" is also exploited, but multi-level sampling has not been considered for multiple sensors;
- if $\boldsymbol{H}_{12}^T\boldsymbol{H}_{12} + \boldsymbol{H}_3^T\boldsymbol{H}_3$ is non-singular, guarantees about sub-gaussian measurements are established in [4], which states that the total number of samples required is dependent on the condition number of $\boldsymbol{H}_{12}^T\boldsymbol{H}_{12} + \boldsymbol{H}_3^T\boldsymbol{H}_3$.

Unfortunately, $\boldsymbol{H}_{12}^T\boldsymbol{H}_{12} + \boldsymbol{H}_3^T\boldsymbol{H}_3$ in our case is singular since it's rank deficient. Thus the available theories about multiple sensors do not apply.

2.2 An Imaginary Sensor

Denote $\boldsymbol{P}_{12} = \boldsymbol{U}_{12}\boldsymbol{\Sigma}_{12}\boldsymbol{V}_{12}^T$ and $\boldsymbol{P}_3 = \boldsymbol{U}_3\boldsymbol{\Sigma}_3\boldsymbol{V}_3^T$ to be the SVDs. Recall the definitions of the sensor profiles $\boldsymbol{H}_{12} = \boldsymbol{I}_3 \otimes \boldsymbol{P}_{12}$ and $\boldsymbol{H}_3 = \boldsymbol{P}_3^T \otimes \boldsymbol{I}_{12}$. It can be shown that

$$\begin{aligned}
&\boldsymbol{H}_{12}^T\boldsymbol{H}_{12} + \boldsymbol{H}_3^T\boldsymbol{H}_3 \\
&= \begin{bmatrix} \boldsymbol{U}_3 \otimes \boldsymbol{V}_{12} & \boldsymbol{U}_3 \otimes \boldsymbol{V}_{12}^\perp & \boldsymbol{U}_3^\perp \otimes \boldsymbol{V}_{12} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_1 & & \\ & \boldsymbol{\Lambda}_2 & \\ & & \boldsymbol{\Lambda}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_3 \otimes \boldsymbol{V}_{12} & \boldsymbol{U}_3 \otimes \boldsymbol{V}_{12}^\perp & \boldsymbol{U}_3^\perp \otimes \boldsymbol{V}_{12} \end{bmatrix}^T,
\end{aligned}$$

where $\boldsymbol{\Lambda}_1 = \boldsymbol{I}_{n_3} \otimes \boldsymbol{\Sigma}_{12}^2 + \boldsymbol{\Sigma}_3^2 \otimes \boldsymbol{I}_{n_1 n_2}$, $\boldsymbol{\Lambda}_2 = \boldsymbol{\Sigma}_3^2 \otimes \boldsymbol{I}_{n_1 n_2}$, $\boldsymbol{\Lambda}_3 = \boldsymbol{I}_{n_3} \otimes \boldsymbol{\Sigma}_{12}^2$. One can show if $\boldsymbol{H}_c = (\boldsymbol{U}_3^\perp)^T \otimes (\boldsymbol{V}_{12}^\perp)^T$, then the combined sensor profiles

$$\boldsymbol{H}_{12}^T\boldsymbol{H}_{12} + \boldsymbol{H}_3^T\boldsymbol{H}_3 + \boldsymbol{H}_c^T\boldsymbol{H}_c$$

is non-singular, and one can tune the weights of $\boldsymbol{H}_{12}$, $\boldsymbol{H}_3$ and $\boldsymbol{H}_c$ to optimize the condition number. Since it's not realistic to measure through $\boldsymbol{H}_c$, and we empirically observe that

$$\|\boldsymbol{H}_c \cdot (\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T) \cdot vec(\boldsymbol{S})\|_2$$

is small, we solve

$$\begin{aligned}
&\min \ \|vec(\boldsymbol{S})\|_1 \\
&\text{s.t.} \begin{bmatrix} \boldsymbol{\Psi}_{12} \cdot \boldsymbol{H}_{12} \\ \boldsymbol{\Psi}_3 \cdot \boldsymbol{H}_3 \\ \boldsymbol{\Psi}_c \cdot \boldsymbol{H}_c \end{bmatrix} \cdot (\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T) \cdot vec(\boldsymbol{S}) \approx \begin{bmatrix} \boldsymbol{\Psi}_{12} \cdot vec(\boldsymbol{Y}) \\ \boldsymbol{\Psi}_3 \cdot vec(\boldsymbol{Z}) \\ \boldsymbol{0} \end{bmatrix}.
\end{aligned} \tag{2}$$

The above formulation

- is theoretically guaranteed (by [4]) to work if $\boldsymbol{\Psi}_{12}$, $\boldsymbol{\Psi}_3$, and $\boldsymbol{\Psi}_c$ are sub-Gaussian measurements;
- introduces imaginary measurement matrix $\boldsymbol{\Psi}_c \cdot \boldsymbol{H}_c$, and therefore is more computationally expensive.

However, as shown in the numerical experiments, the recovery quality of solving (2) is about the same as solving (1). Despite formulation (2) is theoretically guaranteed to work, there is no advantage compared to formulation (1).

[TW: If the $\boldsymbol{0}$ on the right hand side of (2) is replaced by $\boldsymbol{\Psi}_c \cdot \boldsymbol{H}_c \cdot (\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T) \cdot vec(\boldsymbol{S})$ or its noisy version, improved quality of recovery, compared to formulation (1), is observed. But again, it may not be realistic to measure through $\boldsymbol{H}_c$ since it may not correspond to any sensor.]

3 Numerical Experiments

The experiments are conducted with simulated data generated from the Indian Pines dataset. The original data is first reshaped to be 128×128 spatially. Then 4 bands corresponding to blue, green, red, and near infrared spectrum, are generated and considered as the ground truth, so $\mathcal{X} \in \mathbb{R}^{128\times128\times4}$. The HSI $\mathcal{Y} \in \mathbb{R}^{64\times64\times4}$ is generated by convolving with Gaussian blur, downsampling with factor 2, and adding Gaussian noise such that the SNR is about 35. The MSI $\mathcal{Z} \in \mathbb{R}^{128\times128\times1}$ is generated by weighted averaging of the 4 bands, and adding Gaussian noise such that the SNR is about 40. Figure 3 shows the simulated HSI and MSI pair that is used for this first experiment.

Figure 1: Simulated HSI (Left) and MSI (Right) pair. HSI has 4 bands, blue (Top Left), green (Top Right), red (Bottom Left), and near infrared (Bottom Right).

The sparsifying system is chosen to be the Kronecker product of 1D Haar wavelet and 2D Haar wavelet, i.e., $\mathbf{\Phi}_3 \in \mathbb{R}^{4\times4}$ is 1D Haar wavelet with decomposition level 2, and $\mathbf{\Phi}_{12} \in \mathbb{R}^{128\times128}$ is 2D Haar wavelet transform with decomposition level 7. Since HSI and MSI contain about the same amount of information (both of total length 16384), the results reported here are about samples measured equally from HSI and MSI. The combined size of HSI and MSI is *half* the total size of the SRI. We further define the compression ratio ρ as

$$\frac{\text{\# of measurements}}{\text{combined size of HSI and MSI}}.$$

Formulation (1) **with Gaussian Measurements.** The following results are about formulation (1), where $\mathbf{\Phi}_{12}$ and $\mathbf{\Phi}_3$ are i.i.d. Gaussian measurements.

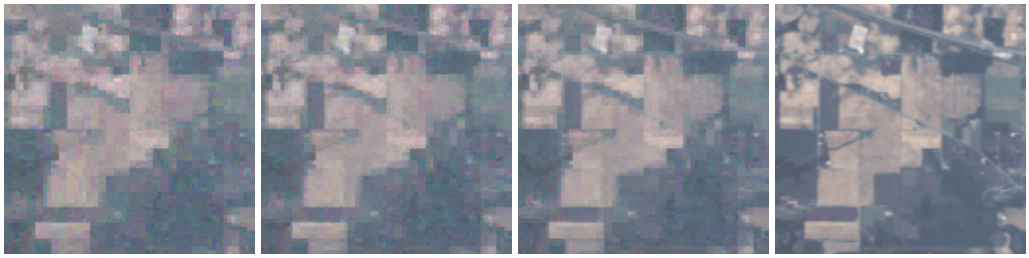

Figure 2: RGB bands of the recovered SRI from various Gaussian measurements and RGB bands of the ground truth. From Left to Right: 8192 measurements ($\rho = 0.25$, PSNR = 27.1507); 12288 measurements ($\rho = 0.375$, PSNR = 28.5207); 16384 measurements ($\rho = 0.5$, PSNR = 29.6215); Ground truth.

Formulation (2) **with Gaussian Measurements.** The following results are about formulation (2), where $\mathbf{\Phi}_{12}$, $\mathbf{\Phi}_3$ and $\mathbf{\Phi}_c$ are i.i.d. Gaussian measurements. The number of rows of $\mathbf{\Phi}_c$ is chosen to be the same as $\mathbf{\Phi}_{12}$ and $\mathbf{\Phi}_3$. One can see that although formulation (2) enjoys theoretical guarantees, adding the imaginary

sensor does not improve the quality of recovery. Also this formulation introduces imaginary measurement matrix $\boldsymbol{\Psi}_c \cdot \boldsymbol{H}_c$, and therefore is more computationally expensive.

Figure 3: RGB bands of the recovered SRI from various Gaussian measurements and RGB bands of the ground truth. From Left to Right: 8192 measurements ($\rho = 0.25$, PSNR = 27.3228); 12288 measurements ($\rho = 0.375$, PSNR = 28.7085); 16384 measurements ($\rho = 0.5$, PSNR = 29.7206); Ground truth.

Formulation (1) **with Hadamard Measurements.** Variable density sampling [7], similar as multi-level sampling [1, 8], can be derived based on local coherences between the measurement matrix (e.g., Hadamard), and the sparsifying matrix. Both matrices are assumed to be unitary. However, in our case, the measurement matrix is

$$\begin{bmatrix} \boldsymbol{M}_{12} \cdot \boldsymbol{H}_{12} \\ \boldsymbol{M}_3 \cdot \boldsymbol{H}_3 \end{bmatrix},$$

where $\boldsymbol{M}_{12}$ and $\boldsymbol{M}_3$ are Kronecker products of 1D and 2D Hadamard matrices of appropriate sizes. Such measurement matrix is generally not unitary. Nevertheless, we can similarly derive the variable density sampling based on local coherences of

$$\begin{bmatrix} \boldsymbol{M}_{12} \cdot \boldsymbol{H}_{12} \\ \boldsymbol{M}_3 \cdot \boldsymbol{H}_3 \end{bmatrix} \cdot (\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T).$$

Note that in this case,

$$\boldsymbol{\Psi}_{12} = \boldsymbol{M}_{12}^{(\Omega_{12},:)}, \quad \boldsymbol{\Psi}_3 = \boldsymbol{M}_3^{(\Omega_3,:)},$$

where Ω_{12} and Ω_3 are the indices of the rows chosen according to local coherences. Improved recovery quality is observed for such measurements, especially when the compression ratio is small.

Figure 4: RGB bands of the recovered SRI from various Hadamard measurements and RGB bands of the ground truth. From Left to Right: 8192 measurements ($\rho = 0.25$, PSNR = 28.9250); 12288 measurements ($\rho = 0.375$, PSNR = 29.4417); 16384 measurements ($\rho = 0.5$, PSNR = 29.9457); Ground truth.

4 Theoretical Guarantees

This section is about theoretical guarantees for the formulation (1). We may only be able to prove for the case when $\boldsymbol{\Psi}_{12}$ and $\boldsymbol{\Psi}_3$ are Gaussian measurements.

4.1 Assumptions

We make two assumptions. The first one is classic in compressed sensing.

Assumption 1. *$vec(\boldsymbol{S})$ can be well-approximated by k nonzero coefficients.*

The next assumption is essential for the fusion from samples to work. Note that

$$\begin{aligned}&vec(\boldsymbol{X})\\ =&\boldsymbol{U}_3 \otimes \boldsymbol{V}_{12} \cdot vec(\boldsymbol{S}_1) + \boldsymbol{U}_3 \otimes \boldsymbol{V}_{12}^{\perp} \cdot vec(\boldsymbol{S}_2) + \boldsymbol{U}_3^{\perp} \otimes \boldsymbol{V}_{12} \cdot vec(\boldsymbol{S}_3) + \boldsymbol{U}_3^{\perp} \otimes \boldsymbol{V}_{12}^{\perp} \cdot vec(\boldsymbol{S}_4),\end{aligned}$$

where

$$\begin{aligned}vec(\boldsymbol{S}_1) =&\boldsymbol{U}_3^T \otimes \boldsymbol{V}_{12}^T \cdot vec(\boldsymbol{X}),\\ vec(\boldsymbol{S}_2) =&\boldsymbol{U}_3^T \otimes (\boldsymbol{V}_{12}^{\perp})^T \cdot vec(\boldsymbol{X}),\\ vec(\boldsymbol{S}_3) =&(\boldsymbol{U}_3^{\perp})^T \otimes \boldsymbol{V}_{12}^T \cdot vec(\boldsymbol{X}),\\ vec(\boldsymbol{S}_4) =&(\boldsymbol{U}_3^{\perp})^T \otimes (\boldsymbol{V}_{12}^{\perp})^T \cdot vec(\boldsymbol{X}).\end{aligned}$$

Recall the sensor profiles

$$\boldsymbol{H}_{12} = \boldsymbol{I}_3 \otimes \boldsymbol{P}_{12} = \boldsymbol{I}_3 \otimes (\boldsymbol{U}_{12}\boldsymbol{\Sigma}_{12}\boldsymbol{V}_{12}^T), \quad \boldsymbol{H}_3 = \boldsymbol{P}_3^T \otimes \boldsymbol{I}_{12} = (\boldsymbol{V}_3\boldsymbol{\Sigma}_3\boldsymbol{U}_3^T) \otimes \boldsymbol{I}_{12},$$

we then have

$$\boldsymbol{H}_{12}\left(\boldsymbol{U}_3^{\perp} \otimes \boldsymbol{V}_{12}^{\perp} \cdot vec(\boldsymbol{S}_4)\right) = \boldsymbol{0}, \quad \boldsymbol{H}_3\left(\boldsymbol{U}_3^{\perp} \otimes \boldsymbol{V}_{12}^{\perp} \cdot vec(\boldsymbol{S}_4)\right) = \boldsymbol{0}.$$

These imply that part of the information of $vec(\boldsymbol{X})$ can never be measured though the two sensors. Thus we cannot expect fusion from samples to produce quality recovery unless the information loss is negligible. Since

$$\begin{aligned}&vec(\boldsymbol{S}_4)\\ =&(\boldsymbol{U}_3^{\perp})^T \otimes (\boldsymbol{V}_{12}^{\perp})^T \cdot vec(\boldsymbol{X})\\ =&(\boldsymbol{\Phi}_3\boldsymbol{U}_3^{\perp})^T \otimes (\boldsymbol{\Phi}_{12}\boldsymbol{V}_{12}^{\perp})^T \cdot vec(\boldsymbol{S}),\end{aligned}$$

$\|\boldsymbol{U}_3^{\perp} \otimes \boldsymbol{V}_{12}^{\perp} \cdot vec(\boldsymbol{S}_4)\|_2 = \|vec(\boldsymbol{S}_4)\|_2$ being negligible motivates the following assumption.

Assumption 2. *Denote $\boldsymbol{s} := vec(\boldsymbol{S})$.*

$$\|(\boldsymbol{\Phi}_3\boldsymbol{U}_3^{\perp})^T \otimes (\boldsymbol{\Phi}_{12}\boldsymbol{V}_{12}^{\perp})^T \cdot \boldsymbol{s}\|_2^2 \leq \xi\|\boldsymbol{s}\|_2^2,$$

where $\xi > 0$ is a small constant.

4.2 Asymmetric RIP

Denote $\boldsymbol{H}_{12}$, $\boldsymbol{H}_3$, and $\boldsymbol{H}_c$ to be the original ones scaled by positive constants a, b and c, respectively. The constants will be specified later. Then there exists $0 < \alpha \leq \beta$ such that

$$\alpha \boldsymbol{I} \preceq \frac{1}{3}(\boldsymbol{H}_{12}^T \boldsymbol{H}_{12} + \boldsymbol{H}_3^T \boldsymbol{H}_3 + \boldsymbol{H}_c^T \boldsymbol{H}_c) \preceq \beta \boldsymbol{I}.$$

For $\boldsymbol{\Psi}_{12} \in \mathbb{R}^{m_{12} \times (n_1 n_2 N_3)}$, $\boldsymbol{\Psi}_3 \in \mathbb{R}^{m_3 \times (N_1 N_2 n_3)}$, and $\boldsymbol{\Psi}_c \in \mathbb{R}^{m_c \times ((N_1 N_2 - n_1 n_2)(N_3 - n_3))}$ being Gaussian measurements, define

$$\begin{aligned}
\boldsymbol{A}_{12} &= \boldsymbol{\Psi}_{12} \boldsymbol{H}_{12} (\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T) \in \mathbb{C}^{m_{12} \times (N_1 N_2 N_3)}, \\
\boldsymbol{A}_3 &= \boldsymbol{\Psi}_3 \boldsymbol{H}_3 (\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T) \in \mathbb{C}^{m_3 \times (N_1 N_2 N_3)}, \\
\boldsymbol{A}_c &= \boldsymbol{\Psi}_c \boldsymbol{H}_c (\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T) \in \mathbb{C}^{m_c \times (N_1 N_2 N_3)},
\end{aligned}$$

and furthermore,

$$\widehat{\boldsymbol{A}} = \begin{bmatrix} \frac{1}{\sqrt{3m_{12}}} \boldsymbol{A}_{12} \\ \frac{1}{\sqrt{3m_3}} \boldsymbol{A}_3 \\ \frac{1}{\sqrt{3m_c}} \boldsymbol{A}_c \end{bmatrix}.$$

Lemma 1. *Suppose $0 < \delta, \epsilon < 1$, and $1 \leq 2k \leq N_1 N_2 N_3$. Define*

$$\Gamma = \frac{1}{\sqrt{\alpha}} \max_j \left\{ \|\boldsymbol{H}_{12}(\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T)\boldsymbol{e}_j\|_2, \|\boldsymbol{H}_3(\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T)\boldsymbol{e}_j\|_2, \|\boldsymbol{H}_c(\boldsymbol{\Phi}_3^T \otimes \boldsymbol{\Phi}_{12}^T)\boldsymbol{e}_j\|_2 \right\}$$

and $L = \ln^2(4k) \ln(2N_1 N_2 N_3) \ln(2(m_{12} + m_3 + \min\{m_{12}, m_3\})) + \ln(2/\epsilon)$. If

$$\min\{m_{12}, m_3\} \geq C \cdot \delta^{-2} \cdot \frac{\beta}{\alpha} \cdot \Gamma^2 \cdot k \cdot L$$

for some universal constant C, then with probability at least $1 - \epsilon$,

$$(\alpha - \delta\alpha) \cdot \|\boldsymbol{w}\|_2^2 \leq \|\widehat{\boldsymbol{A}}\boldsymbol{w}\|_2^2 \leq (\beta + \delta\alpha) \cdot \|\boldsymbol{w}\|_2^2, \quad \forall \boldsymbol{w} \in \Sigma_{2k},$$

where $\Sigma_k = \left\{ \boldsymbol{w} \in \mathbb{C}^{N_1 N_2 N_3 \times 1} \mid \|\boldsymbol{w}\|_0 \leq k \right\}$.

Proof. Let $m_c = \min\{m_{12}, m_3\}$, the proof follows directly from [4, Theorem 2.8]. □

4.3 Recovery Guarantee

Let $t = \sqrt{\frac{2}{\beta + \alpha}}$. Then $\widehat{\boldsymbol{A}}' = t\widehat{\boldsymbol{A}}$ satisfies the standard RIP (of order $2k$), i.e.,

$$(1 - \delta') \cdot \|\boldsymbol{w}\|_2^2 \leq \|\widehat{\boldsymbol{A}}'\boldsymbol{w}\|_2^2 \leq (1 + \delta') \cdot \|\boldsymbol{w}\|_2^2, \quad \forall \boldsymbol{w} \in \Sigma_{2k}$$

for $\delta' = 1 - \frac{2(1-\delta)\alpha}{\beta + \alpha}$. The following l^2 robust null space property follows immediately.

Lemma 2. *If the $2k$ restricted isometry constant of $\widehat{\boldsymbol{A}}'$ satisfies $\delta' < \frac{4}{\sqrt{41}}$, then for any $\boldsymbol{w}$ and set Δ of cardinality no more k,*

$$\|\boldsymbol{w}_\Delta\|_2 \leq \frac{\rho}{\sqrt{k}} \|\boldsymbol{w}_{\Delta^c}\|_1 + \tau \|\widehat{\boldsymbol{A}}'\boldsymbol{w}\|_2 \tag{3}$$

holds for

$$\rho = \frac{\delta'}{\sqrt{1 - (\delta')^2 - \delta'/4}} \in (0, 1), \quad \tau = \frac{\sqrt{1 + \delta'}}{\sqrt{1 - (\delta')^2 - \delta'/4}} > 0.$$

Proof. The proof can be found in [5, Theorem 6.13]. □

[TW: The constraint $\delta' < \frac{4}{\sqrt{41}}$ requires $\frac{\beta}{\alpha} < \frac{2(1-\delta)}{1-\frac{4}{\sqrt{41}}} - 1 < 4.33$.]

As a consequence, the following result holds.

Lemma 3. *Suppose $\widehat{\boldsymbol{A}}'$ satisfies the l^2 robust null space property with $\rho \in (0,1)$ and $\tau > 0$. Fix $\boldsymbol{w}$ and let Δ be the locations of the largest k elements in magnitude of $\boldsymbol{w}$. Then*

$$\|\boldsymbol{w}\|_2 \leq \left(\rho + \sqrt{\rho} + \frac{1}{2}\right) \frac{\|\boldsymbol{w}_{\Delta^c}\|_1}{\sqrt{k}} + \frac{3\tau}{2} \|\widehat{\boldsymbol{A}}'\boldsymbol{w}\|_2.$$

Proof. The result corresponds to the case $l = 1$ in [2, Lemma 6.1]. □

[TW: The weights can be changed to $\left(\rho + \sqrt{\rho} + \frac{1}{2\lambda}\right)$ and $\left(1 + \frac{\lambda}{2}\right)\tau$.]

The l^2 robust null space property also implies the l^1 robust null space property. Indeed, if (3) holds, then

$$\|\boldsymbol{w}_\Delta\|_1 \leq \rho \|\boldsymbol{w}_{\Delta^c}\|_1 + (\tau\sqrt{k}) \|\widehat{\boldsymbol{A}}'\boldsymbol{w}\|_2.$$

The l^1 robust null space property implies the following result.

Lemma 4. *If $\widehat{\boldsymbol{A}}'$ satisfies the l^1 robust null space property with $\rho \in (0,1)$ and $\tau\sqrt{k} > 0$, then*

$$\|\boldsymbol{w}^{(1)} - \boldsymbol{w}^{(2)}\|_1 \leq \frac{1+\rho}{1-\rho} \left(\|\boldsymbol{w}^{(1)}\|_1 - \|\boldsymbol{w}^{(2)}\|_1 + 2\|\boldsymbol{w}^{(2)}_{\Delta^c}\|_1\right) + \frac{2\tau\sqrt{k}}{1-\rho} \|\widehat{\boldsymbol{A}}'(\boldsymbol{w}^{(1)} - \boldsymbol{w}^{(2)})\|_2$$

holds for any $\boldsymbol{w}^{(1)}$, $\boldsymbol{w}^{(2)}$, and Δ of cardinality no more than k.

Proof. The proof can be found in [5, Theorem 4.20]. □

[TW: $\frac{2}{1-\rho}$ can be improved to $\frac{2-\rho}{1-\rho}$.]

For the ease of presentation, define

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{\sqrt{3m_{12}}} \boldsymbol{A}_{12} \\ \frac{1}{\sqrt{3m_3}} \boldsymbol{A}_3 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} \boldsymbol{\Psi}_{12} \cdot vec(\boldsymbol{Y}) \\ \boldsymbol{\Psi}_3 \cdot vec(\boldsymbol{Z}) \end{bmatrix},$$

$\boldsymbol{A}' = t\boldsymbol{A}$, $\boldsymbol{b}' = t\boldsymbol{b}$ and $\eta' = t\eta$. Furthermore, define $\boldsymbol{A}'_{\boldsymbol{c}} = \frac{t}{\sqrt{3m_c}} \boldsymbol{A}_c$. Our main result in the following.

Theorem 1. *Suppose the underlying $\boldsymbol{s}$ satisfies $\|\boldsymbol{A}'\boldsymbol{s} - \boldsymbol{b}'\|_2 \leq \eta'$. Then any solution $\boldsymbol{f}$ to*

$$\begin{aligned} &\min_{\boldsymbol{f}'} \ \|\boldsymbol{f}'\|_1 \\ &\text{s.t. } \|\boldsymbol{A}'\boldsymbol{f}' - \boldsymbol{b}'\|_2 \leq \eta' \end{aligned}$$

satisfy

$$\|\boldsymbol{f} - \boldsymbol{s}\|_2 \leq \cdots$$

Proof. For any Δ of cardinality k,

$$\begin{aligned}
\|\boldsymbol{f}-\boldsymbol{s}\|_1 \leq & \frac{1+\rho}{1-\rho}\left(\|\boldsymbol{f}\|_1-\|\boldsymbol{s}\|_1+2\|\boldsymbol{s}_{\Delta^c}\|_1\right)+\frac{2\tau\sqrt{k}}{1-\rho}\|\widehat{\boldsymbol{A}}'(\boldsymbol{f}-\boldsymbol{s})\|_2 \\
\leq & \frac{1+\rho}{1-\rho}\left(\|\boldsymbol{f}\|_1-\|\boldsymbol{s}\|_1+2\sigma_k(\boldsymbol{s})\right)+\frac{2\tau\sqrt{k}}{1-\rho}\|\widehat{\boldsymbol{A}}'(\boldsymbol{f}-\boldsymbol{s})\|_2 \\
\leq & \frac{2+2\rho}{1-\rho}\sigma_k(\boldsymbol{s})+\frac{2\tau\sqrt{k}}{1-\rho}\|\widehat{\boldsymbol{A}}'(\boldsymbol{f}-\boldsymbol{s})\|_2 \\
= & \frac{2+2\rho}{1-\rho}\sigma_k(\boldsymbol{s})+\frac{2\tau\sqrt{k}}{1-\rho}\sqrt{\|\boldsymbol{A}'(\boldsymbol{f}-\boldsymbol{s})\|_2^2+\|\boldsymbol{A}'_c(\boldsymbol{f}-\boldsymbol{s})\|_2^2} \\
\leq & \frac{2+2\rho}{1-\rho}\sigma_k(\boldsymbol{s})+\frac{2\tau\sqrt{k}}{1-\rho}\left(\|\boldsymbol{A}'(\boldsymbol{f}-\boldsymbol{s})\|_2+\|\boldsymbol{A}'_c(\boldsymbol{f}-\boldsymbol{s})\|_2\right) \\
\leq & \frac{2+2\rho}{1-\rho}\sigma_k(\boldsymbol{s})+\frac{2\tau\sqrt{k}}{1-\rho}\left(2\eta'+\|\boldsymbol{A}'_c(\boldsymbol{f}-\boldsymbol{s})\|_2\right)
\end{aligned}$$

where the first inequality is due to Lemma 4; in the second inequality, $\sigma_k(\boldsymbol{s})$ is the residue of the best k-sparse approximation to $\boldsymbol{s}$; the third inequality is because $\boldsymbol{f}$ is a minimizer of the l^1 norm; the fifth inequality is due to $\|\boldsymbol{A}'(\boldsymbol{f}-\boldsymbol{s})\|_2=\|\boldsymbol{A}'\boldsymbol{f}-\boldsymbol{b}'-(\boldsymbol{A}'\boldsymbol{s}-\boldsymbol{b}')\|_2\leq\|\boldsymbol{A}'\boldsymbol{f}-\boldsymbol{b}'\|_2+\|\boldsymbol{A}'\boldsymbol{s}-\boldsymbol{b}'\|_2$.

$$\begin{aligned}
\|\boldsymbol{f}-\boldsymbol{s}\|_2 \leq & \left(\rho+\sqrt{\rho}+\frac{1}{2}\right)\frac{\|(\boldsymbol{f}-\boldsymbol{s})_{\Delta^c}\|_1}{\sqrt{k}}+\frac{3\tau}{2}\|\widehat{\boldsymbol{A}}'(\boldsymbol{f}-\boldsymbol{s})\|_2 \\
\leq & \left(\rho+\sqrt{\rho}+\frac{1}{2}\right)\frac{\|\boldsymbol{f}-\boldsymbol{s}\|_1}{\sqrt{k}}+\frac{3\tau}{2}\|\widehat{\boldsymbol{A}}'(\boldsymbol{f}-\boldsymbol{s})\|_2 \\
\leq & \left(\rho+\sqrt{\rho}+\frac{1}{2}\right)\frac{\|\boldsymbol{f}-\boldsymbol{s}\|_1}{\sqrt{k}}+\frac{3\tau}{2}\left(2\eta'+\|\boldsymbol{A}'_c(\boldsymbol{f}-\boldsymbol{s})\|_2\right) \\
\leq & \left(\rho+\sqrt{\rho}+\frac{1}{2}\right)\frac{2+2\rho}{1-\rho}\cdot\frac{\sigma_k(\boldsymbol{s})}{\sqrt{k}}+\left(\rho+\sqrt{\rho}+\frac{1}{2}\right)\frac{2\tau}{1-\rho}\left(2\eta'+\|\boldsymbol{A}'_c(\boldsymbol{f}-\boldsymbol{s})\|_2\right) \\
& +\frac{3\tau}{2}\left(2\eta'+\|\boldsymbol{A}'_c(\boldsymbol{f}-\boldsymbol{s})\|_2\right) \\
= & C_1(\delta')\cdot\frac{\sigma_k(\boldsymbol{s})}{\sqrt{k}}+2C_2(\delta')\cdot\eta'+C_2(\delta')\cdot\|\boldsymbol{A}'_c(\boldsymbol{f}-\boldsymbol{s})\|_2
\end{aligned}$$

where the first inequality follows from applying Lemma 3 by setting $\boldsymbol{w}=\boldsymbol{f}-\boldsymbol{s}$, and Δ corresponds to the largest k elements in magnitude of $\boldsymbol{w}$; in the equality, $C_1(\delta')=\left(\rho+\sqrt{\rho}+\frac{1}{2}\right)\frac{2+2\rho}{1-\rho}$, $C_2(\delta')=\left(\rho+\sqrt{\rho}+\frac{1}{2}\right)\frac{2\tau}{1-\rho}+\frac{3\tau}{2}$ since ρ, τ are defined only using δ'.

$$
\begin{aligned}
\|\boldsymbol{A}_c'(\boldsymbol{f}-\boldsymbol{s})\|_2 =& \frac{t}{\sqrt{3m_c}}\|\boldsymbol{A}_c(\boldsymbol{f}-\boldsymbol{s})\|_2 \\
=& \frac{t}{\sqrt{3m_c}}\|\boldsymbol{\Psi}_c\boldsymbol{H}_c(\boldsymbol{\Phi}_3^T\otimes\boldsymbol{\Phi}_{12}^T)(\boldsymbol{f}-\boldsymbol{s})\|_2 \\
\leq& \sqrt{\frac{2}{\beta+\alpha}}\left\|\frac{1}{\sqrt{3m_c}}\boldsymbol{\Psi}_c\right\|_2\|\boldsymbol{H}_c\|_2\|\boldsymbol{f}-\boldsymbol{s}\|_2 \\
\lesssim& \sqrt{\frac{2\alpha}{3(\beta+\alpha)}}\|\boldsymbol{f}-\boldsymbol{s}\|_2 \lesssim \sqrt{\frac{1-\delta'}{3}}\|\boldsymbol{f}-\boldsymbol{s}\|_2
\end{aligned}
$$

[TW: We would like $C_3(\delta') = C_2(\delta')\sqrt{\frac{1-\delta'}{3}}$ to be less than 1 to derive the final bound on $\|\boldsymbol{f}-\boldsymbol{s}\|_2$. However, $C_3(\delta') > 1.4$ even for small δ'. The bound in [5, Theorem 4.25] is not better.]

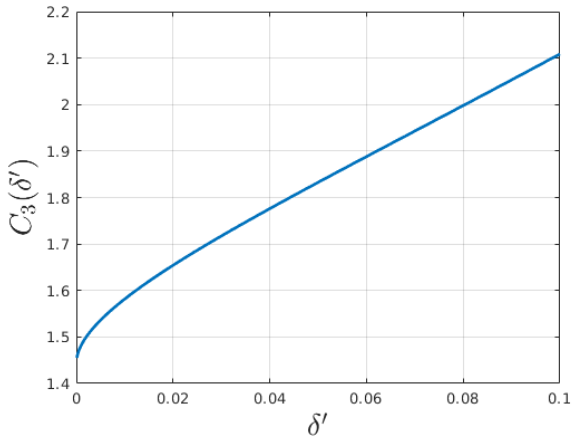

Figure 5: Function value of $C_3(\delta')$ when $\delta' \in (0, 0.1)$.

□

References

[1] Ben Adcock, Anders C Hansen, Clarice Poon, and Bogdan Roman. Breaking the coherence barrier: A new theory for compressed sensing. In *Forum of Mathematics, Sigma*, volume 5. Cambridge University Press, 2017.

[2] Alexander Bastounis and Anders C Hansen. On the absence of uniform recovery in many real-world applications of compressed sensing and the restricted isometry property and nullspace property in levels. *SIAM Journal on Imaging Sciences*, 10(1):335–371, 2017.

[3] Il Yong Chun and Ben Adcock. Compressed sensing and parallel acquisition. *IEEE Transactions on Information Theory*, 63(8):4860–4882, 2017.

[4] Il Yong Chun and Ben Adcock. Uniform recovery from subgaussian multi-sensor measurements. *Applied and Computational Harmonic Analysis*, 2018.

[5] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013.

[6] Tatiana Gelvez and Henry Arguello. Spectral image fusion from compressive projections using total-variation and low-rank regularizations. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1985–1989. IEEE, 2018.

[7] Felix Krahmer and Rachel Ward. Stable and robust sampling strategies for compressive imaging. *IEEE transactions on image processing*, 23(2):612–622, 2013.

[8] Chen Li and Ben Adcock. Compressed sensing with local structure: uniform recovery guarantees for the sparsity in levels class. *Applied and Computational Harmonic Analysis*, 2017.

[9] Amirafshar Moshtaghpour, José M Bioucas-Dias, and Laurent Jacques. Compressive single-pixel Fourier transform imaging using structured illumination. *arXiv preprint arXiv:1810.13200*, 2018.