

Rank-Order Graphs for Signal Recovery and Its Application to Compressive Spectral Imaging

Juan F. Florez-Ospina, Daniel L. Lau, *Senior Member, IEEE*, Dominique Guillot, Kenneth Barner, *Fellow, IEEE*, and Gonzalo R. Arce, *Fellow, IEEE*

Abstract—Compressed sensing (CS) enables the recovery of signals from undersampled linear measurements by exploiting the assumption that most signals are (approximately) sparse with respect to a given dictionary. In doing so, an ℓ_1 -norm regularization problem is typically solved where, additionally, side-information signals can be used to improve reconstructions. In this paper, we depart from the ℓ_1 -norm regularization problem and, instead, consider smoothness with respect to path graphs defined by rank orderings inferred from the side information. To this end, we develop the concept of rank-order graph, and reconstruct the signal of interest by solving a greatly preferred ℓ_2 -norm regularization problem. We develop theoretical results along with illustrative examples. The proposed framework is adapted to tackle the problem of compressive spectral imaging (CSI) with panchromatic side information. Our numerical and real experiments reveal the advantages of our method over traditional CSI approaches. To conclude, we discuss how rank-order graphs can be extended to a variety of signal reconstruction problems and how they can be exploited for adaptive CSI systems.

Index Terms—Signal Recovery from Undersampled Measurements, Compressive Spectral Imaging, Graph Laplacian Regularization, Rank-Order Path Graphs

1 INTRODUCTION

SIGNAL RECOVERY from an incomplete or limited number of linear observations is a fundamental inverse problem [1], which appears in a myriad of applications—from medical imaging [2] and seismic exploration [3] to baggage and cargo inspection in homeland security [4]. Since the problem is inherently underdetermined with infinitely many solutions, an underlying signal of interest can only be reconstructed provided that extraneous criteria of the signal is available. The additional information often specifies regularity (or structural) constraints that narrow down the available solutions, leaving only those with desirable properties, and thereby allowing accurate signal estimates to be found.

The premise that most signals can be well-approximated by sparse signals is one of the most popular forms of prior knowledge used to derive regularity constraints [1, 5]. In this setting, the signal of interest is assumed to have a sparse representation with respect to a given dictionary, and the solution to the inverse problem is given by the signal with the sparsest representation that complies with the measurements. Sparsity has motivated the study of model-based approaches such as block-sparsity [6] and union-of-subspace models [1], which in certain scenarios can outperform sparsity-based reconstructions.

In the last decade, numerous approaches that go beyond sparsity have been proposed. Some exploit state-of-the-art denoisers [7] while others use Gaussian mixture models [8]

to regularize the inverse problem. More recently, generative models have been used for the same purpose [9]. Such models are capable of learning mappings between a low dimensional representation space and a space of signals with desirable properties. Although this approach has great flexibility to tackle a variety of problems, it comes at a high price in terms of sample complexity, which might be difficult to have in scientific applications where there is little to no training data [10].

In certain imaging inverse problems, approaches such as Newmann networks [10] and deep image priors [11] decrease sample complexity by leveraging either the sensing matrix during training or regularity constraints inferred from the observations themselves. Most (partially) data-driven approaches, despite their remarkable performance, do so at the expense of interpretability [12]. That is to say, the underlying mechanisms that drive reconstructions are not explicit, and therefore it may be difficult to put those approaches in a more formal framework.

In this paper, we would like to promote the use of rank-ordering information to regularize the problem of signal recovery from undersampled linear measurements. To do so, we introduce the concept of *rank-order graph*, and assume that the best solution to the inverse problem is the signal that is smoothest with respect to a given rank-order graph. Rank-order graphs arise from merging a collection of path graphs, which are constructed on a given ranking of the nodes. Owing to the sparse nature of path graphs, rank-order graphs have highly sparse edge structure, which allows them to be easily integrated in large scale optimization problems.

Our approach stems from the key observation that most natural signals admit a smooth rearrangement of their values. Consider, for example, the random step signal and its nondecreasing rearrangement in Fig. 1. In the rearranged

Juan F. Florez-Ospina, Kenneth Barner, and Gonzalo R. Arce are with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, 19716 USA e-mail: {jflopez, alejopm, barner, arce}@udel.edu.
Daniel L. Lau is with the Department of Electrical Engineering, University of Kentucky, Lexington, KY, 40506 USA e-mail: dllau@uky.edu
Dominique Guillot is with the Dept. of Mathematical Sciences, University of Delaware, Newark, DE, 19716 USA e-mail: dguillot@udel.edu

domain, we may exploit the simplicity of smoothness priors, that is a graph-based smoothness prior, to tackle the aforementioned signal reconstruction problem. In fact, it has been shown that several image restoration tasks can be simplified to one-dimensional signal processing tasks by finding suitable rearrangements of image patches [13], or even data compression in a database can be increased by rearranging the order in which data is stored [14].

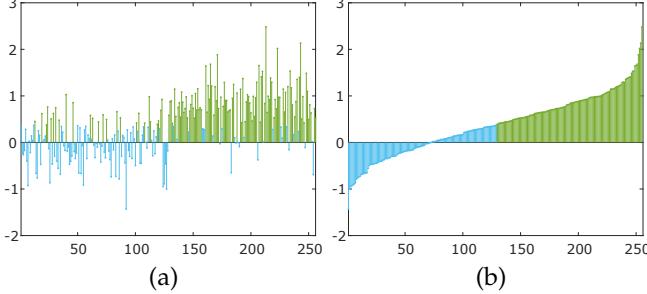


Fig. 1. (a) Random step signal and (b) its non-decreasing rearrangement. Blue and green colors indicate lower and higher ranked samples.

Unlike sparse representation approaches that use the ℓ_1 -norm to find feasible solutions, the proposed approach uses the ℓ_2 -norm, which leads to a solution that can be written in closed-form, and does not require explicit definition of a dictionary. Like sparsity-based models that require the definition of suitable sparsifying bases, the success of our approach depends on the availability of suitable rank-ordering information. Obviously, the best ordering information can be extracted from the signal of interest itself, so it might seem restrictive at first; however, there are numerous applications where such ordering information can be approximated by a side information signal, which is statistically related to the signal of interest and can be acquired inexpensively. As an alternative, we could connect with the agenda of learning-based priors, and try to learn the best ordering information from readily available information. This is, however, outside of the scope of the paper.

In this article, we showcase the suitability of rank-order graphs as a means to accurately reconstruct spectral images from compressive measurements. In imaging spectroscopy, we would like to characterize a scene of interest by sensing large amounts of spatial information across a multitude of frequency bands. Since the efficient sensing of all these bands is challenging, compressive spectral imaging (CSI) was developed to reconstruct a hyperspectral image from a spatio-spectrally coded snapshot of the scene [15]. But a compressive snapshot is usually insufficient to allow for the high quality reconstruction of complex spectral scenes, so we can often rely on broadband panchromatic detectors that provide side information to improve reconstruction quality [16, 17, 18, 19]. This poses an ideal scenario where the ordering properties of the side information can be used to construct the rank-orderings, which can be used to regularize the CSI inverse problem.

2 SIGNAL RECOVERY USING GRAPHS

As aforementioned, we consider the problem of reconstructing a signal of interest $\mathbf{x} \in \mathbb{R}^n$ from a number of (noisy) undersampled linear measurements $\mathbf{y} \in \mathbb{R}^m$, obeying

$$\mathbf{y} = \mathbf{Ax} + \mathbf{n}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a sensing matrix with fewer rows than columns, i.e., $m < n$, and \mathbf{n} is an additive noise term. This problem is known to have infinitely many solutions, that is there are many signals $\hat{\mathbf{x}}$ that satisfy $\mathbf{A}\hat{\mathbf{x}} = \mathbf{Ax}$, and therefore regularity constraints are needed to solve the problem [20]. Traditionally, one would assume that \mathbf{x} belongs to the set of k -sparse signals $\mathcal{M}_k = \{\mathbf{x} \in \mathbb{R}^n, \|\Psi^T \mathbf{x}\|_0 \leq k\}$, where k represents the sparsity level of a signal on a given dictionary Ψ . Stable recovery of \mathbf{x} would then follow from solving the problem [1, 20]:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\Psi^T \mathbf{x}\|_1 \text{ s.t. } \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \epsilon \quad (2)$$

provided that \mathbf{A} satisfies the null-space property when the noise level $\epsilon = 0$ or similarly the restricted isometry property (RIP) when $\epsilon > 0$.

Our approach to solving (1) instead suggests that the best solution to the underdetermined system of linear equations is the signal that is smoothest with respect to a given graph. A graph $G = (V, E, w)$ is an ordered triple, where $V = \{1, 2, \dots, n\}$ is a vertex set, $E \subset V \times V$ is an edge set, and $w : E \mapsto [0, \infty)$ is a non-negative weight function. An edge is an unordered pair (i, j) of vertices $i, j \in V$. We only consider graphs without loops, i.e., $(i, i) \notin E$ for any $i \in V$. And a weight function w satisfies that $w(i, j) = w(j, i) > 0$ for $(i, j) \in E$, and $w(i, j) = 0$ for $(i, j) \notin E$. When w is unspecified, we assume the graph is unweighted, or equivalently that $w(i, j) = 1$ for $(i, j) \in E$ and zero otherwise.

The smoothest signal with respect to a given graph G can then be reconstructed by solving the problem:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{L}_G \mathbf{x}\|_2 \text{ s.t. } \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \epsilon, \quad (3)$$

where $\mathbf{L}_G \in \mathbb{R}^{n \times n}$ is the graph Laplacian¹ of G with entries i, j given by

$$(\mathbf{L}_G)_{ij} = \begin{cases} \sum_{j=1}^n w(i, j), & i = j \\ -w(i, j), & (i, j) \in E \\ 0, & (i, j) \notin E. \end{cases} \quad (4)$$

Note that the program (3), as opposed to the program (2), imposes a less stringent condition on the sensing matrix \mathbf{A} , which may be useful in applications where structured rather than dense matrices represent the sensing process. In particular, it is required that the null space of \mathbf{L}_G be approximately orthogonal to the null space of \mathbf{A} for stable reconstructions as will be elaborated Sec. 2.2.

2.1 Graph smoothness and sparsity

To see how a measure of smoothness with respect to a graph G and sparsity with respect to a dictionary Ψ may be related, consider the following proposition:

Proposition 1. Suppose $\mathbf{x} \in \mathbb{R}^n$ is a signal on the vertex set V of a connected graph G , that is x_i stores the value of \mathbf{x} at the vertex $i \in V$. Let $\psi_j \in \mathbb{R}^n, i = 1, \dots, n$ be the eigenvectors of the graph Laplacian \mathbf{L}_G , that is $\mathbf{L}_G \psi_j = \lambda_j \psi_j$ where $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$. Then the measure

¹An important characteristic of the graph Laplacian is that it is symmetric positive semidefinite, thus its eigenvalues are real and nonnegative.

of smoothness $\|\mathbf{L}_G \mathbf{x}\|_2$ of the signal \mathbf{x} with respect to G satisfies

$$\|\mathbf{W} \Psi^T \mathbf{x}\|_1 \leq \|\mathbf{L}_G \mathbf{x}\|_2 \quad (5)$$

where $\Psi = (\psi_1, \dots, \psi_n)$, $\mathbf{W} \in \mathbb{R}^{n \times n}$ is a diagonal matrix such that $W_{ii} = \lambda_i^2 / \sum_{j=1}^n \lambda_j^2$.

Proof. See Appendix A \square

Proposition 1 highlights the fact that the function $\mathbf{x} \mapsto \|\mathbf{L}_G \mathbf{x}\|_2$ may act as a surrogate of the sparsity-promoting function $\mathbf{x} \mapsto \|\mathbf{W} \Psi^T \mathbf{x}\|_1$. In particular, by either minimizing $\mathbf{x} \mapsto \|\mathbf{W} \Psi^T \mathbf{x}\|_1$ or $\mathbf{x} \mapsto \|\mathbf{L}_G \mathbf{x}\|_2$ over the feasible set $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{y}\}$, we can find signal estimates with similar characteristics. To clarify, as indicated in the proof of Proposition 1, both $\|\mathbf{L}_G \mathbf{x}\|_2^2$ and $\|\mathbf{W} \Psi^T \mathbf{x}\|_1$ can be written as $\sum_{i=1}^n \lambda_i^2 |\langle \psi_j, \mathbf{x} \rangle|^2$ and $C \sum_{i=1}^n \lambda_i^2 |\langle \psi_j, \mathbf{x} \rangle|$, with $C > 0$ respectively. Since $\lambda_i \leq \lambda_{i+1}$ for all $i \in V$, both measures of smoothness with respect to G penalize signals with energy concentrated at the higher eigenvectors more than signals with energy concentrated at the lower eigenvectors and therefore promote signal estimates, which are smooth in the vertex domain or equivalently low-pass spectrum in the frequency domain.

However, note that the dictionary Ψ , unlike the graph Laplacian \mathbf{L}_G , has in general an explicitly-dense structure, which may be difficult to compute and store in high dimensional signal recovery applications. Therefore, unless the dictionary Ψ has an explicit functional form, it is perhaps most useful to use \mathbf{L}_G , which one typically designs to have sparse structure. In most cases the dictionary Ψ or the graph G are not known *a priori*, and must be designed based on our prior knowledge of the signal of interest. A dictionary can be designed based on pre-constructed bases (such as wavelets), or examples of the signal of interest [21, 22]. Comparably, a graph can be constructed or learned from signals observed at the nodes of the graph [23]. We present our approach to construct suitable graphs with desirable properties in Sec. 3.

2.2 Stability and Uniqueness

To address the uniqueness and stability of the estimate $\hat{\mathbf{x}}$ of $\mathbf{x} \in \mathbb{R}^n$ obtained using the program (3), we consider the case where $\epsilon = 0$, that is $\mathbf{A}\mathbf{x} = \mathbf{y}$, and consider the following class of signals:

$$\mathcal{M}_\varepsilon = \{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{L}_G \mathbf{x}\|_2 \leq \varepsilon\} \quad (6)$$

where the parameter $\varepsilon > 0$ controls the smoothness of the signals in \mathcal{M}_ε with respect to G . Given that $\mathbf{x} \in \mathcal{M}_\varepsilon$, an important question to ask is: what properties should the sensing matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ satisfy in order for \mathbf{x} to be uniquely identified from compressed measurements $\mathbf{y} = \mathbf{A}\mathbf{x}$ where $m < n$? To address this question, consider the following theorem.

Theorem 1. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{M}_\varepsilon$ be such that $\mathbf{x}_1 \neq \mathbf{x}_2$. Then the measurements $\mathbf{A}\mathbf{x}_1, \mathbf{A}\mathbf{x}_2$ satisfy that $\mathbf{A}\mathbf{x}_1 \neq \mathbf{A}\mathbf{x}_2$ if and only if the null space of \mathbf{A} does not intersect with any vectors in $\mathcal{M}_{2\varepsilon}$ except at the zero vector, i.e., $\mathcal{N}_\mathbf{A} \cap \mathcal{M}_{2\varepsilon} = \{\mathbf{0}\}$.

Proof. See Appendix B. \square

This theorem provides a general characterization of the type of sensing matrices \mathbf{A} that would permit a complete identification of a signal $\mathbf{x} \in \mathcal{M}_\varepsilon$ from compressed measurements of the form $\mathbf{y} = \mathbf{A}\mathbf{x}$. In particular, the null space of \mathbf{A} has to be orthogonal to the set $\mathcal{M}_{2\varepsilon}$.

Remark 1. Note that $\mathcal{N}_{\mathbf{L}_G} \subseteq \mathcal{M}_{2\varepsilon}$. Given that $\mathcal{N}_\mathbf{A} \cap \mathcal{M}_{2\varepsilon} = \{\mathbf{0}\}$, we obtain that $\mathcal{N}_\mathbf{A} \cap \mathcal{N}_{\mathbf{L}_G} = \{\mathbf{0}\}$. Since $\mathcal{N}_\mathbf{A}$ and $\mathcal{N}_{\mathbf{L}_G}$ are complementary subspaces, the dimension of their sum $\dim(\mathcal{N}_\mathbf{A} + \mathcal{N}_{\mathbf{L}_G})$ can be written as $\dim \mathcal{N}_\mathbf{A} + \dim \mathcal{N}_{\mathbf{L}_G}$, and strictly upper bounded by n . On the other hand, the theorem of conservation of dimension for matrices [24] gives us that $\dim \mathcal{N}_\mathbf{A} + \dim \mathcal{R}_\mathbf{A} = n$, so since $\dim \mathcal{R}_\mathbf{A} \leq m$, we get the lower bound $n - m \leq \dim \mathcal{N}_\mathbf{A}$. Therefore the number of measurements m must obey

$$\dim \mathcal{N}_{\mathbf{L}_G} < m. \quad (7)$$

This relates the number of connected components of the graph G , i.e., $\dim \mathcal{N}_{\mathbf{L}_G}$, to the number of measurements.

We now proceed to characterize the estimation error, which may be used to study the accuracy and stability of the signal estimates. To do so, consider first the following Lemma.

Lemma 1 (Restricted strong convexity). Assume $\mathcal{N}_\mathbf{A} \cap \mathcal{M}_{2\varepsilon} = \{\mathbf{0}\}$. Then there exist $\gamma > 0$ such that the objective function $\|\mathbf{L}_G \mathbf{v}\|_2$ satisfies

$$\gamma \leq \frac{\|\mathbf{L}_G \mathbf{v}\|_2}{\|\mathbf{v}\|_2} \quad (8)$$

for all nonzero $\mathbf{v} \in \mathcal{N}_\mathbf{A}$.

Proof. See Appendix C \square

This lemma plays an important role in upper-bounding the estimation error as will be elaborated shortly. It shows that the objective function behaves as a strongly convex function when restricted to the null space of the sensing matrix \mathbf{A} . In the context of signal recovery, this means that the function $\mathbf{x} \mapsto \|\mathbf{L}_G \mathbf{x}\|_2$ has a unique global optimum on the set $\{\mathbf{x} \in \mathbb{R}^n, \mathbf{A}\mathbf{x} = \mathbf{y}\}$. It is important to note that the condition $\mathcal{N}_\mathbf{A} \cap \mathcal{M}_{2\varepsilon} = \{\mathbf{0}\}$ can be further relaxed to $\mathcal{N}_\mathbf{A} \cap \mathcal{M}_\varepsilon = \{\mathbf{0}\}$, in which case the error bound won't be as tight, but we are still able to produce stable signal estimates.

Theorem 2. Let $\mathcal{N}_\mathbf{A} \cap \mathcal{M}_{2\varepsilon} = \{\mathbf{0}\}$. Then for any signal of interest $\mathbf{x} \in \mathcal{M}_\varepsilon$, for $\epsilon = 0$, the estimate $\hat{\mathbf{x}}$ obtained via the program (3) has approximation error

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2\varepsilon(\mathbf{x})}{\gamma} \leq \frac{2\varepsilon}{\gamma} \quad (9)$$

where $\varepsilon(\mathbf{x}) = \|\mathbf{L}_G \mathbf{x}\|_2$, and $\gamma > 0$ is the restricted strong convexity constant with respect to $\mathcal{N}_\mathbf{A}$.

Proof. See Appendix D \square

This theorem exhibits two important factors that influence the approximation error. First, the term $\varepsilon(\mathbf{x})$ is proportional to the error, thus we could potentially aim at better approximations by better inferring graphs G on which our signal of interest is known to be smooth. A good starting point is always that the graph has only one connected component so that Eq. (7) is satisfied. Second, the boundedness of the error

depends essentially on the constant γ , which results from the interaction between the sensing matrix \mathbf{A} and the graph Laplacian \mathbf{L}_G . We will elaborate on how this interaction arises and how it can be used to design adaptive sensing schemes for compressive spectral imaging in the discussion section.

3 RANK-ORDER GRAPHS

We now design a class of graphs, referred to as *rank-order graphs*, that can be used to find an accurate estimates of a signal from measurements obeying (1) using the program (3).

Let \mathcal{S}_n denote the set of permutations (or rankings) of V where a permutation, $\sigma \in \mathcal{S}_n$, may be written as a linear ordering $\sigma = \sigma_1\sigma_2\dots\sigma_{n-1}\sigma_n$ where $\sigma_i \in V$ is listed exactly once. For instance, the set \mathcal{S}_3 of permutations of $V = \{1, 2, 3\}$ consists of 123, 132, 213, 231, 312, and 321. Alternatively, a permutation may be written as a bijective map $\sigma : V \mapsto V$ such that $\sigma(1) = \sigma_1$, $\sigma(2) = \sigma_2, \dots, \sigma(n-1) = \sigma_{n-1}, \sigma(n) = \sigma_n$. For instance, $\sigma \in \mathcal{S}_3$ can be given by $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$.

A (*rank-order*) *path graph* on the vertex set $V = \{1, 2, \dots, n\}$ is a graph $G = (V, E)$ with edge set $E \subset V \times V$ given by

$$E = \{(\sigma_1, \sigma_2), (\sigma_2, \sigma_3), \dots, (\sigma_{n-1}, \sigma_n)\}$$

where σ is a permutation (or ranking) of V . Figure 2 shows two pictorial representations of rank-order path graphs on $V = \{1, 2, 3, 4, 5\}$, which are associated with the permutations 12345 and 13542.

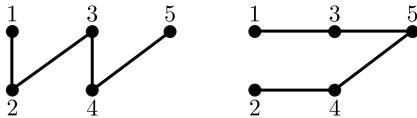


Fig. 2. Illustration of path graphs \mathcal{P}_{σ_1} on the left and \mathcal{P}_{σ_2} on the right.

A notable characteristic of a (*rank-order*) path graph G with associated permutation $\sigma \in \mathcal{S}_n$ is that the eigenvectors Ψ of its graph Laplacian \mathbf{L}_G obey $\Psi = \mathbf{P}^T \mathbf{U}$, where \mathbf{U} is the $n \times n$ discrete cosine transform matrix (DCT-2) [25], and \mathbf{P} is a $n \times n$ binary matrix such that $\mathbf{P} = \sum_{j=1}^n \mathbf{e}_j \mathbf{e}_{\sigma_j}^T$, where \mathbf{e}_j is the j -th standard unitary vector. Although this property may turn useful in the design of reconstruction algorithms [26], in our application we exploit primarily the associated graph Laplacian, which has a significantly sparse structure.

3.1 Rank-order Path Graphs for Signal Reconstruction

In our context, to identify the graph G , we begin with the simplifying assumption that G belongs to the set of unweighted path graphs \mathcal{L}_n of length n on the vertex set $V = \{1, \dots, n\}$. In this manner, the objective function in (3) can be decomposed into two quadratic terms.

Proposition 2. Let $\mathbf{x} \in \mathbb{R}^n$ be a signal on V , and let $G \in \mathcal{L}_n$ be a path graph. Then, for any $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{L}_G \mathbf{x}\|_2$ can be expressed as

$$\sqrt{2} \left(\mathbf{x}^T \mathbf{L}_G \mathbf{x} - \sum_{i=2}^{n-1} (x_{\sigma_i} - x_{\sigma_{i-1}})(x_{\sigma_{i+1}} - x_{\sigma_i}) \right)^{1/2} \quad (10)$$

where $\sigma \in \mathcal{S}_n$ dictates the edge set of G in a path graph. Consequently,

$$\mathbf{x}^T \mathbf{L}_G \mathbf{x} - \sum_{i=2}^{n-1} (x_{\sigma_i} - x_{\sigma_{i-1}})(x_{\sigma_{i+1}} - x_{\sigma_i}) \geq 0 \quad (11)$$

Proof. See the supplementary materials. \square

The proposition exhibits the fact that if we would like a signal of interest \mathbf{x} to have minimal Laplacian semi-norm $\|\mathbf{L}_G \mathbf{x}\|_2$ as required by (3), the path graph G should minimize the difference between the quadratic terms in (11) over \mathcal{L}_n . To find such a graph may require extensive computation. Nonetheless, by minimizing the quadratic form $\mathbf{x}^T \mathbf{L}_G \mathbf{x}$ in (11), one can find explicitly a suitable rank-order path graph as illustrated in the following theorem.

Theorem 3. Let $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ and let $\sigma^* \in \mathcal{S}_n$ be a permutation such that

$$x_{\sigma_1^*} \leq x_{\sigma_2^*} \leq \dots \leq x_{\sigma_n^*}. \quad (12)$$

Then the function

$$F(G) = \mathbf{x}^T \mathbf{L}_G \mathbf{x} \quad (13)$$

achieves a global minimum on \mathcal{L}_n at the graph G^* with edges $\{(\sigma_i^*, \sigma_{i+1}^*) : i = 1, \dots, n-1\}$.

Proof. See Appendix E. \square

When the graph G is restricted to the set of path graphs \mathcal{L}_n , the best path graph is constructed on the permutation or rank ordering of the vertex set V , which rearranges the elements of the signal of interest \mathbf{x} in ascending order. As one might expect, it is not possible to have access to the permutation σ^* without knowing the signal of interest \mathbf{x} itself. However, there is a variety of applications, such as compressive spectral imaging with side information, where we can approximate such a permutation and still obtain remarkable results.

3.2 Side Information

We now address the problem of signal recovery from undersampled measurements using a rank-order path graph and introduce an auxiliary signal \mathbf{z} , which may be readily acquired by some inexpensive sensing mechanism and may be used to approximate the ideal permutation σ^* . Our interest here is to recover a signal of interest $\mathbf{x} \in \mathbb{R}^n$ from measurements $\mathbf{y} \in \mathbb{R}^m$ obeying

$$\mathbf{y} = \mathbf{A} \mathbf{x} \quad (14)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a known matrix with $m < n$. Clearly, since the nullspace of \mathbf{A} is nontrivial, we cannot simply invert \mathbf{A} to find \mathbf{x} and, therefore, need additional constraints. Assume then that in addition to \mathbf{y} , we have access to a side signal \mathbf{z} such that

$$z_{\hat{\sigma}_1} \leq \dots \leq z_{\hat{\sigma}_n} \Rightarrow x_{\hat{\sigma}_1} \leq \dots \leq x_{\hat{\sigma}_n} \quad (15)$$

where $\hat{\sigma} \in \mathcal{S}_n$ is a permutation of $\{1, 2, \dots, n\}$ that rearranges the entries of \mathbf{z} in ascending order. Then the following constraint set can be defined

$$\mathcal{M} = \{\mathbf{x} \in \mathbb{R}^n : x_{\hat{\sigma}_1} \leq x_{\hat{\sigma}_2} \leq \dots \leq x_{\hat{\sigma}_n}\}, \quad (16)$$

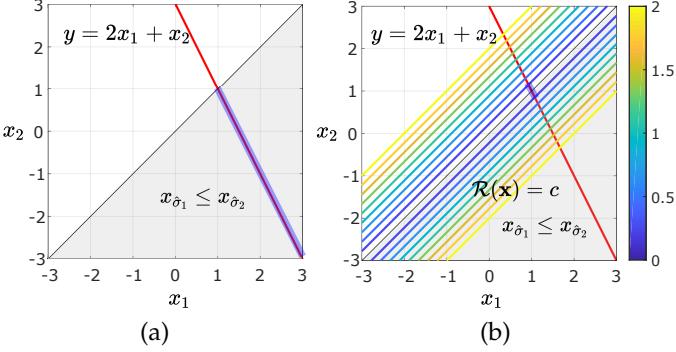


Fig. 3. Geometric illustration of underdetermined system of linear equations subject to ordering constraints. (a) The shaded region illustrates the ordering relationship between the components of \mathbf{x} in (16). (b) The contour lines of the objective function, which favor signals whose induced order statistics are smooth.

obtaining a new feasible set of signals $\mathcal{M} \cap \{\mathbf{x} \in \mathbb{R}^m : \mathbf{Ax} = \mathbf{y}\}$. A natural question to ask is: can we identify a unique estimate of \mathbf{x} on such feasible set? The answer is no, and we still need additional assumptions on the nature of \mathbf{x} . To understand why, consider the following example.

Example 1. Consider the recovery of a two-dimensional signal $\mathbf{x} = (x_1, x_2)^T$ from a scalar linear measurement $y = \mathbf{Ax}$ with $\mathbf{A} = (2, 1)$. For the sake of simplicity, assume that we have access to a signal $\mathbf{z} \in \mathbb{R}^2$, which is rank correlated to \mathbf{x} , i.e., $z_{\hat{\sigma}_1} \leq z_{\hat{\sigma}_2} \Rightarrow \mathcal{M} = \{\mathbf{x} \in \mathbb{R}^2, x_{\hat{\sigma}_1} \leq x_{\hat{\sigma}_2}\}$. Note that the set $\{\mathbf{x} \in \mathbb{R}^2, \mathbf{Ax} = \mathbf{y}\} \cap \{\mathbf{x} \in \mathbb{R}^2, x_{\hat{\sigma}_1} \leq x_{\hat{\sigma}_2}\}$, highlighted in blue in Fig. 3(a), does not allow us to distinguish an unique estimate of \mathbf{x} . Observe however that under the assumption that \mathbf{x} is smooth with respect to the path graph $G \in \mathcal{L}_2$ with edge set $\{(\hat{\sigma}_1, \hat{\sigma}_2)\}$, it is possible to identify unique estimates of \mathbf{x} . Note that the contour lines of $R(\mathbf{x}) = \|\mathbf{L}_{\hat{G}}\mathbf{x}\|_2$, shown in Fig. 3(b), illustrate the fact that signals with smooth reordered entries are favored or preferred.

Therefore, given (14) and (16), and the fact that the signal of interest is smooth with respect to the path graph $\hat{G} \in \mathcal{L}_n$ with edge set $\{(\hat{\sigma}_1, \hat{\sigma}_2), (\hat{\sigma}_2, \hat{\sigma}_3), \dots, ((\hat{\sigma}_{n-1}, \hat{\sigma}_n))\}$, an accurate estimate of \mathbf{x} can be obtained by solving

$$\min_{\mathbf{x} \in \mathcal{M}} \|\mathbf{L}_{\hat{G}}\mathbf{x}\|_2 \text{ s.t. } \mathbf{Ax} = \mathbf{y}, \quad (17)$$

where $\hat{G} = G^*$ is constructed as in Theorem 3. Needless to say that in practice we do not have access to G^* but to an approximate graph \hat{G} . We now elaborate on that case.

3.2.1 The case where \mathbf{x} and \mathbf{z} are not perfectly rank-correlated

In the case where (15) does not hold, i.e. \mathbf{x} and \mathbf{z} are not perfectly rank-correlated, we can still obtain accurate estimates of \mathbf{x} by relying only on the smoothness of \mathbf{x} with respect to \hat{G} . The interest is thus not in finding estimates of \mathbf{x} that strictly belong in \mathcal{M} since $\mathbf{x} \notin \mathcal{M}$. Instead, we look for signal estimates which are proximate to \mathcal{M} in the sense that those signal estimates are the smoothest with respect to the underlying path graph \hat{G} . To clarify, note that \mathcal{M} can be

rewritten as $\mathcal{M} = \{\mathbf{x} \in \mathbb{R}^n, \mathbf{M}\mathbf{x} \leq \mathbf{0}\}$ where the entries of \mathbf{M} are given by

$$(\mathbf{M})_{ij} = \begin{cases} 1, & \hat{\sigma}_j = \hat{\sigma}_i \\ -1, & \hat{\sigma}_j = \hat{\sigma}_{i+1} \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

for $(i, j) \in \{1, \dots, n-1\} \times \{1, \dots, n\}$. Consequently, the graph Laplacian of \hat{G} can be expressed² in terms of \mathbf{M} as

$$\mathbf{L}_{\hat{G}} = \mathbf{M}^T \mathbf{M}. \quad (19)$$

This exhibits that estimates of \mathbf{x} that minimize the function $\mathbf{x} \mapsto \|\mathbf{L}_{\hat{G}}\mathbf{x}\|_2$ over the set $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{y}\}$ are proximate to the null space of \mathcal{M} . More formally, in practice, we solve the problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{L}_{\hat{G}}\mathbf{x}\|_2 \text{ s.t. } \mathbf{Ax} = \mathbf{y}, \quad (20)$$

which is a relaxation of (17). As stated in Theorem 2, the estimation error $\|\hat{\mathbf{x}} - \mathbf{x}\|_2$ is upperbounded by the value $\|\mathbf{L}_{\hat{G}}\mathbf{x}\|_2$, given that the sensing matrix \mathbf{A} satisfies restricted strong convexity. In this setting, one may also be interested in characterizing how well the graph \hat{G} approximates G^* . In the following theorem, we measure the goodness of the approximation in terms of the Euclidean proximity of the projections $\mathbf{L}_{\hat{G}}\mathbf{x}$ and $\mathbf{L}_{G^*}\mathbf{x}$ of the signal of interest \mathbf{x} .

Theorem 4. Let $\mathbf{x} = (x_1, \dots, x_n)^T$ be a scalar signal on V , and let $G^* \in \mathcal{L}_n$ be a path graph constructed as in Theorem 3. For any $G \in \mathcal{L}_n$, the prediction error $\|(\mathbf{L}_G - \mathbf{L}_{G^*})\mathbf{x}\|_2$ obeys

$$\|(\mathbf{L}_G - \mathbf{L}_{G^*})\mathbf{x}\|_2 \leq 8\|\mathbf{L}_G^{1/2}\mathbf{x}\|_2. \quad (21)$$

As an immediate consequence, one has

$$\|(\mathbf{L}_G - \mathbf{L}_{G^*})\mathbf{x}\|_2 \leq 8|E| \max_{1 \leq i \leq n-1} |x_{\sigma_i} - x_{\sigma_{i+1}}| \quad (22)$$

where $\sigma \in \mathcal{S}_n$ is the permutation associated to G , and $|E|$ is the number of elements of the edge set of G .

Proof. See Appendix F □

This theorem shows that the path-graph approximation error is controlled by the measure of smoothness $\|\mathbf{L}_G^{1/2}\mathbf{x}\|_2$ of \mathbf{x} with respect to G , and in particular by the number of vertices $|V|$, or edges $|E| = |V| - 1$, and the proximity of adjacent values $|x_{\sigma_i} - x_{\sigma_{i+1}}|$ of the reordered signal of interest \mathbf{x} .

3.3 Local Rank-Order Path Graphs

We now elaborate on the graph topology that will be most relevant to our application in compressive spectral imaging, and more generally in settings where the interest is in the recovery of discrete time signals and images. As suggested by Theorem (4), one can diminish the path-graph approximation error (22) by decreasing the number of ranking units (the number of vertices), and the maximum absolute deviation among samples with adjacent rank orders. To do so, one can then use local rather than global rank-order relationships. This can be naturally done in the case where the signals of interest are discrete-time signals or

²The matrix \mathbf{M}^T is the incidence matrix of a directed graph.

images since both temporal and spatial proximity can be respectively used to define localities where samples exhibit small sample variation.

Accordingly, let $S \subset V$ be a user-defined set of vertices (or sampling set), and let $\Omega_k \subset V$ be a subset of vertices indexed by $k \in S$ such that $V = \bigcup_{k \in S} \Omega_k$. For instance, when \mathbf{x} is a discrete-time signal, we can define $\Omega_k = \{v \in V : |v - k| \leq r\}$ where r is the radius of Ω_k around k . Now, let $\Omega_k = \{s_i^k\}_{i=1}^{|\Omega_k|}$ be an ordering of the elements of Ω_k such that the entries of the side signal \mathbf{z} at Ω_k satisfy

$$z_{s_1^k} \leq z_{s_2^k} \leq \dots \leq z_{s_{|\Omega_k|}^k}. \quad (23)$$

We can then define a local ordering model as

$$\mathcal{M}_k = \{\mathbf{x} \in \mathbb{R}^n : x_{s_1^k} \leq \dots \leq x_{s_{|\Omega_k|}^k}\}. \quad (24)$$

In this manner, each local ordering model \mathcal{M}_k is associated to a local rank-order graph G_k , that is a graph with vertex set $V = \{1, \dots, n\}$, edge set

$$E_k = \{(s_i^k, s_{i+1}^k) : i = 1, \dots, |\Omega_k|\}, \quad (25)$$

and weight function $w^{(k)} : E_k \mapsto \{0, 1\}$ such that

$$w^{(k)}(i, j) = \frac{\mathbb{1}_{E_k}(i, j)}{\sum_{k=1}^{|S|} \mathbb{1}_{E_k}(i, j)} \quad (26)$$

where $\mathbb{1}_{E_k}$ denotes the indicator function of the edge set E_k .

Let $p_1, \dots, p_{|S|}$ be user-defined positive weights. Then the recovery $\mathbf{x} \in \mathbb{R}^n$ from measurements $\mathbf{y} \in \mathbb{R}^m$ can be performed by minimizing the following objective function

$$\mathbf{x} \mapsto \|\mathbf{L}_{\hat{G}} \mathbf{x}\|_2 \quad (27)$$

over the set of feasible signals $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{y}\}$ where \hat{G} is a rank-order graph with graph Laplacian $\mathbf{L}_{\hat{G}}$ given by

$$\mathbf{L}_{\hat{G}} = \sum_{k=1}^{|S|} \sqrt{p_k} \mathbf{L}_{G_k}. \quad (28)$$

To understand the measure of smoothness $\|\mathbf{L}_{\hat{G}} \mathbf{x}\|_2$ of \mathbf{x} with respect to the graph \hat{G} , and study how it induces regularity on a signal estimate, consider the following proposition.

Proposition 3. Let $\mathbf{x} \in \mathbb{R}^n$ be a scalar signal on $V = \{1, \dots, n\}$, and let $\{G_k\}_{k=1}^K$ be a collection of rank-order graphs on V constructed as above. Then the measure of smoothness $\|\mathbf{L}_{\hat{G}} \mathbf{x}\|_2$ of \mathbf{x} with respect to \hat{G} satisfies that

$$\begin{aligned} \|\mathbf{L}_{\hat{G}} \mathbf{x}\|_2 &\leq \sum_{k=1}^{|S|} p_k \|\mathbf{L}_{G_k} \mathbf{x}\|_2 + \\ &\sum_{j=1}^{|S|} \sum_{k=1, k \neq j}^{|S|} (p_j p_k)^{1/2} |\langle \mathbf{L}_{G_j} \mathbf{x}, \mathbf{L}_{G_k} \mathbf{x} \rangle|^{1/2} \end{aligned} \quad (29)$$

This proposition exhibits that estimates of \mathbf{x} obtained by minimizing $\|\mathbf{L}_{\hat{G}} \mathbf{x}\|_2$ must be locally smooth with respect to the local rank-order graphs $\{G_k\}_{k=1}^{|S|}$, as suggested by the first term of the r.h.s. in (29). Furthermore, note that the estimates must be smooth across adjacent signal segments as suggested by the minimization of the cross-terms $|\langle \mathbf{L}_{G_j} \mathbf{x}, \mathbf{L}_{G_k} \mathbf{x} \rangle|$. However, notice that $|\langle \mathbf{L}_{G_j} \mathbf{x}, \mathbf{L}_{G_k} \mathbf{x} \rangle| = 0$

when $\Omega_j \cap \Omega_k = \emptyset$. Therefore, to guarantee smoothness across segments there must always be some overlap among adjacent localities, i.e., if Ω_j and Ω_k are adjacent, one must have $\Omega_j \cap \Omega_k \neq \emptyset$.

By adapting the weights $\{p_k\}_{k=1}^{|S|}$ to the local statistics of \mathbf{z} , one may add additional flexibility to rank-order graphs. For example, one might want to loosen regularity on a region Ω_k where there are jump discontinuities by setting p_k slightly smaller than surrounding regions. In this paper, however, our interest is to explore solely the role of rank-ordering information, and therefore these weights are set to one, $p_1 = p_2 = \dots = p_{|S|} = 1$.

Alternatively, the graph \hat{G} with graph Laplacian $\mathbf{L}_{\hat{G}}$ given by (28) can be constructed using the following graph operation.

Graph operation 1 (Edge Union of Graphs). Let $V = \{1, 2, \dots, n\}$ be a set of vertices, $p_1, \dots, p_{|S|}$ be a collection of nonnegative weights, and $\{G_k\}_{k=1}^{|S|}$ be a finite set of graphs with vertex set V , and edge set E_k . Then the edge union between the graphs $\{G_k\}_{k=1}^{|S|}$ is given by:

$$G = (V, \bigcup_k E_k, w) \quad (30)$$

where the weight function $w : \bigcup_k E_k \mapsto [0, \infty)$ is defined by

$$w(i, j) = \sum_k \sqrt{p_k} w^{(k)}(i, j) \quad (31)$$

Example 2. Set $V = \{1, 2, \dots, 6\}$, $S = \{1, 4\}$, $r = 2$, and $p_1 = p_4 = 1$. One can obtain sets $\Omega_1 = \{1, 2, 3\}$ and $\Omega_4 = \{2, 3, 4, 5, 6\}$ and define local ordering models \mathcal{M}_1 and \mathcal{M}_4 with orderings $\Omega_1 = \{s_i^1\}_{i=1}^3$ and $\Omega_4 = \{s_i^4\}_{i=1}^5$ given by

$$\begin{pmatrix} s_1^1 & s_2^1 & s_3^1 \\ 1 & 2 & 3 \end{pmatrix} \text{ and } \begin{pmatrix} s_1^4 & s_2^4 & s_3^4 & s_4^4 & s_5^4 \\ 2 & 3 & 4 & 6 & 5 \end{pmatrix}, \quad (32)$$

respectively. Then the rank-order graph \hat{G} can be obtained by applying the graph operation 1 to G_1 and G_4 , which are the graphs associated with the local ordering models \mathcal{M}_1 and \mathcal{M}_4 . A pictorial representation of G_1 , G_4 , and \hat{G} is shown in Fig. 4.

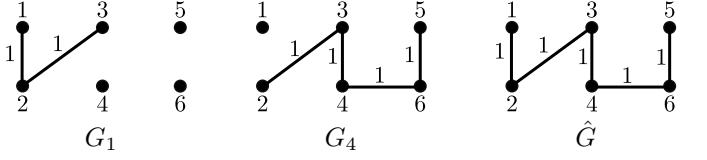


Fig. 4. Graph representation of the ordering models in Example 2. From left to right, it can be observed the graphs G_1 , G_4 , and \hat{G} .

3.4 Illustrative Examples

Consider the signals $\mathbf{x} \in \mathbb{R}^n$, in Fig. 5, and let $\mathbf{y} \in \mathbb{R}^m$ denote a set of noisy undersampled measurements such that $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$ where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a rank-deficient sensing matrix with entries $(\mathbf{A})_{ij} \sim \mathcal{N}(0, 1)$, and \mathbf{n} is a Gaussian-noise vector with entries $n_i \sim \mathcal{N}(0, 0.05)$. In addition, assume that we have access to a side signal $\mathbf{z} \in \mathbb{R}^n$ perfectly rank-correlated to \mathbf{x} , that is \mathbf{x} and \mathbf{z} satisfy (15), and let \hat{G}_1 be a rank-order path graph constructed as in Theorem 3, \hat{G}_2 be

a rank-order graph obtained by applying Graph operation 1 to a collection of local rank-order graphs $\{G_k\}_{k=1}^{|S|}$ where $G_k = (V, E_k, w^{(k)})$ is constructed as in (25). The smoothest estimates of \mathbf{x} with respect to the underlying graphs \hat{G}_1 and \hat{G}_2 can thus be found by minimizing $\|\mathbf{L}_{\hat{G}_1} \mathbf{x}\|_2$ and $\|\mathbf{L}_{\hat{G}_2} \mathbf{x}\|_2$ respectively over the set $\{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \varepsilon\}$ where $\varepsilon = 0.05\sqrt{m} + \sqrt{2m}$, which is a consequence of the error being a chi-square random variable as elaborated in [20].

For $n = 512$, $m = 179$, graph parameters $r = 50$, $S = \{1 + r(k-1)\}_{k=1}^{n-r}$, and $p_k = 1$, reconstructed and error signals are shown in Fig. 6. Observe that when both local and global ordering information about the signal of interest is available, high-PSNR reconstructions can be obtained by seeking smoothness with respect to the proposed graphs \hat{G}_1 and \hat{G}_2 . Here the rank-order graph \hat{G}_2 constructed from local ordering models is outperformed by the rank-order path graph \hat{G}_1 constructed from global ordering information of the signal of interest.

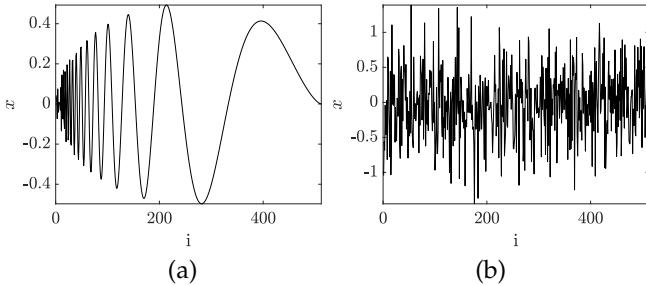


Fig. 5. (a) Doppler and (b) Random signals to be recovered from under-sampled measurements using smoothness with respect to rank-order graphs.

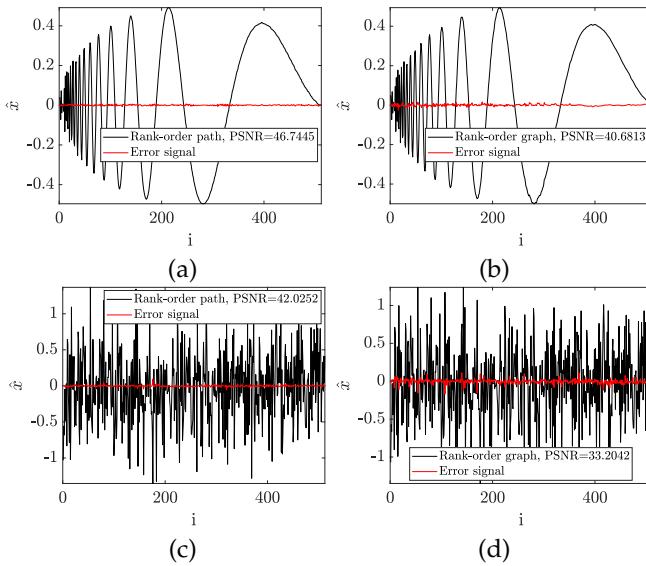


Fig. 6. Reconstructed and error signals using smoothness with respect to the rank-order path graph \hat{G}_1 and the rank-order graph \hat{G}_2 . (a) and (b) display the reconstructions of the Doppler signal using \hat{G}_1 and \hat{G}_2 respectively. (c) and (d) display reconstructions of the Random signal using \hat{G}_1 and \hat{G}_2 respectively. (a),(b),(c), and (d) obtain PSNR values of 45.34, 40.66, 40.08, and 35.25 dBs respectively.

However, as mentioned in Sec. 3, when imperfect ordering information is available, \hat{G}_2 may be preferred over \hat{G}_1 . To clarify, given a graph G , define its perturbed graph Laplacian as $\mathbf{P}^T \mathbf{L}_G \mathbf{P}$ where $\mathbf{P} \in \{0, 1\}^{n \times n}$ is a permutation

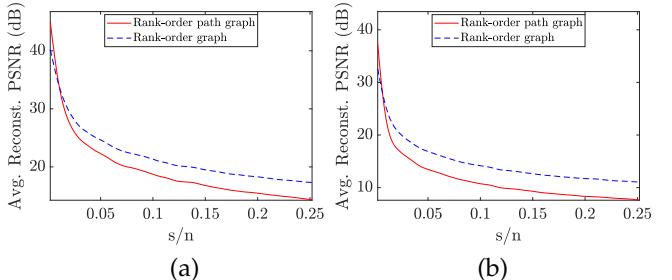


Fig. 7. Average PSNR value of the reconstructed (a) Doppler and (b) Random signals using the rank-order graphs \hat{G}_1 and \hat{G}_2 subjected to s out of n random vertex-ordering permutations.

matrix such that $\mathbf{P} = \mathbf{P}_{v_1} \mathbf{P}_{v_2} \dots \mathbf{P}_{v_s}$ where the collection of permutation matrices $\mathbf{P}_{v_1}, \dots, \mathbf{P}_{v_s}$ is indexed by a set of s vertices selected uniformly at random, and \mathbf{P}_{v_i} inverts the ordering of the vertices v_{i-1} and v_{i+1} . We reconstruct \mathbf{x} from \mathbf{y} provided the perturbed graph Laplacians associated to \hat{G}_1 and \hat{G}_2 . The average reconstruction performance for the perturbed graphs \hat{G}_1 and \hat{G}_2 against fraction of random vertex-ordering permutations over 50 trials is shown in Fig. 7. Note that the rank-order graph \hat{G}_2 is more robust against such perturbations than the single rank-order path graph \hat{G}_1 .

4 COMPRESSIVE HYPERSPECTRAL IMAGING WITH PANCHROMATIC SIDE INFORMATION

To demonstrate a practical application for solving underdetermined systems of linear equations via rank-order graph smoothness, we consider the dual-camera compressive spectral imaging (DCCSI) system, depicted in Fig. 8, which aims at the reconstruction of a spectral image of a scene of interest with L spectral bands $\mathbf{F} = (\mathbf{F}_1, \dots, \mathbf{F}_L) \in \mathbb{R}^{n_1 \times n_2 \times L}$ from a single spatio-spectrally coded image $\mathbf{G} \in \mathbb{R}^{n_1 \times n_2 + L-1}$, captured by the FPA1, obeying [27]:

$$\mathbf{g} \approx \mathbf{H}_{\text{CASSI}} \mathbf{f} \quad (33)$$

and a co-registered panchromatic image $\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$ of the scene, captured by the FPA2, obeying

$$\mathbf{z} \approx \mathbf{R}_{\text{PAN}} \mathbf{f} \quad (34)$$

where \mathbf{g} , \mathbf{z} , \mathbf{f} are vectorized versions of \mathbf{G} , \mathbf{Z} and \mathbf{F} .

The $\mathbf{H}_{\text{CASSI}}$ and \mathbf{R}_{PAN} represent the forward model of single-disperser coded-aperture snapshot spectral imaging (CASSI) system in the horizontal arm and the forward model of the side-information camera in the vertical arm. Alternatively, the (i, j) -entry of \mathbf{G} can be expressed as $G_{i,j} \approx \sum_{l=1}^L H_{i-l+1,j} F_{i-l+1,j,l}$ where $H_{i-l+1,j} = \Delta^2 \Delta_l s_1(\lambda_l) \tau(\Delta(i-l+1), \Delta_j)$ represent the entries of $\mathbf{H}_{\text{CASSI}}$, and Δ , Δ_l , $s_1(\lambda_l)$ and $\tau(\cdot)$ denote the detector's pitch size, the l -th band's bandwidth, the detector's spectral sensitivity, and the coded aperture transmittance respectively [15, 27]. Similarly, the (i, j) entry of \mathbf{Z} can be expressed as $Z_{i,j} = \sum_{l=1}^L s_2(\lambda_l) F_{i,j,l}$ where $s_2(\lambda_l)$ denotes the spectral sensitivity of the FPA2. Since the set of linear equations (33) and (34) is underdetermined, we cannot directly invert the forward operators to find an accurate reconstruction of the spectral image \mathbf{f} . In the next subsections, we present some traditional approaches and our approach to solve the problem.

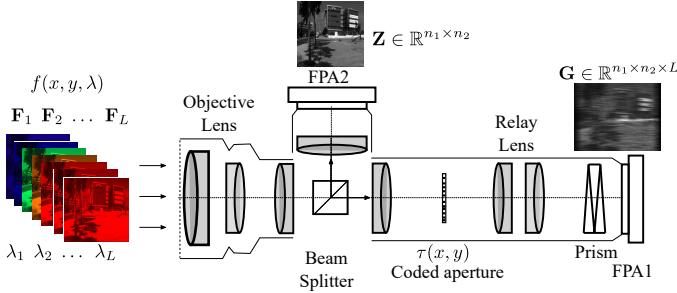


Fig. 8. Schematic of single disperser CASSI system with side information. In the horizontal arm, the objective lens images the spectral scene which is relayed through the beam-splitter onto the coded aperture mask by means of the auxiliary lens. Next, the spatially coded image is relayed through an Amici prism, which outputs an spatio-spectrally coded image, onto the FPA1. Similarly, the FPA2 captures a panchromatic image of the scene.

4.1 Reconstruction via Sparsity

Assume that the spectral scene \mathbf{f} has minimal total-variation (TV) norm $\|\mathbf{f}\|_{\text{TV}}$ over the feasible set. Then one can reconstruct \mathbf{f} by solving the problem

$$\min_{\mathbf{f} \in \mathbb{R}^{n_1 n_2 L}} \|\mathbf{f}\|_{\text{TV}} \quad \text{s. t.} \quad \left\| \begin{pmatrix} \mathbf{g} \\ \mathbf{z} \end{pmatrix} - \begin{pmatrix} \mathbf{H}_{\text{CASSI}} \\ \mathbf{R}_{\text{PAN}} \end{pmatrix} \mathbf{f} \right\|_2 \leq \epsilon \quad (35)$$

where $\epsilon \geq 0$ is the noise level. Here, the TV norm has to be extended to the vectorial case, i.e., the sum of the TV norm of each spectral band [27, 28, 29], or more formally $\|\mathbf{f}\|_{\text{TV}} = \sum_{l=1}^L \sum_{i,j} \|\nabla(\mathbf{F}_l)_{ij}\|_2$ where $\nabla \approx (\frac{d}{dx}, \frac{d}{dy})^T$. Some other useful extensions are considered in [30].

Similarly, one can also assume that the spectral scene \mathbf{f} is the sparsest on a given dictionary Ψ , that is \mathbf{f} has minimal l_1 -norm $\|\Psi^T \mathbf{f}\|_1$ over the feasible set. As a consequence, the reconstruction problem can be cast as:

$$\min_{\mathbf{f} \in \mathbb{R}^{n_1 n_2 L}} \|\Psi^T \mathbf{f}\|_1 \quad \text{s. t.} \quad \left\| \begin{pmatrix} \mathbf{g} \\ \mathbf{z} \end{pmatrix} - \begin{pmatrix} \mathbf{H}_{\text{CASSI}} \\ \mathbf{R}_{\text{PAN}} \end{pmatrix} \mathbf{f} \right\|_2 \leq \epsilon. \quad (36)$$

In our experiments, we adopt the Kronecker product of the 2D-Wavelet-Symlet8 transform and the 1D-discrete cosine transform $\Psi = \Psi_{\text{W2D}} \otimes \Psi_{\text{DCT}}$ because it has proven a suitable basis for CSI reconstruction [15, 18].

4.2 Reconstruction via Graphs

In this setting, we assume that each spectral band is (sufficiently) smooth with respect to a given graph \hat{G} . To capture this notion of smoothness, let $\mathbf{f} = (\mathbf{f}_1^T, \dots, \mathbf{f}_L^T)^T$, where $\mathbf{f}_l = \text{vec}(\mathbf{F}_l)$ is the l -th spectral band in vectorized form, and measure the smoothness of the spectral image as follows:

$$\|\mathbf{L}_{\hat{G}} \mathbf{f}\|_2 = \sqrt{\sum_{l=1}^L w_l \|\mathbf{L}_{\hat{G}} \mathbf{f}_l\|_2^2}, \quad (37)$$

where w_1, \dots, w_L are positive weights, and $\mathbf{L}_{\hat{G}} = \text{diag}(w_1, \dots, w_L) \otimes \mathbf{L}_{\hat{G}}$ is a graph Laplacian matrix with L connected components, i.e., $\dim(\mathcal{N}_{\mathbf{L}_{\hat{G}}}) = L + 1$. A natural concern with this smoothness model may be the implicit independence assumption among adjacent bands, which goes counter the fact that most hyperspectral images have naturally correlated bands. However, in this setting, the independence assumption among channels would work

provided that the CASSI measurements \mathbf{G} encode well the nature of such correlation. Therefore, our approach can be formulated as:

$$\min_{\mathbf{f} \in \mathbb{R}^{n_1 n_2 L}} \sqrt{\sum_{l=1}^L w_l \|\mathbf{L}_{\hat{G}} \mathbf{f}_l\|_2^2} \quad \text{s. t.} \quad \|\mathbf{g} - \mathbf{H}_{\text{CASSI}} \mathbf{f}\|_2 \leq \epsilon \quad (38)$$

where the graph \hat{G} is constructed using the ordering properties of \mathbf{Z} as explained in Sec. 3.3. In particular, the construction procedure is now explained.

4.2.1 Rank-order graph construction

Let $\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$ be a panchromatic image on the vertex set $V = \{1, 2, \dots, n_1 n_2\}$, that is each $k \in V$ is associated to a pixel position $(k_y, k_x) \in \{1, \dots, n_1\} \times \{1, \dots, n_2\}$. We define a sampling set $S \subset V$, and a patch radius r . For simplicity, let Z_k denote the pixel value (or entry) of \mathbf{Z} at the pixel coordinate (k_y, k_x) . For each $k \in S$, perform

- 1) Define a square neighborhood³ Ω_k around k such that $\Omega_k = \{i \in V : \max\{|i_y - k_y|, |i_x - k_x|\} \leq r\}$.
- 2) Find an ordering $\{s_i^k\}_{i=1}^{|\Omega_k|}$ of Ω_k such that the values of \mathbf{Z} at Ω_k are placed in ascending order as in (23).
- 3) Construct a local rank-order graph $G_k = (V, E_k)$ on V with edge set given by (25).

Use Graph operation 1 to construct a final rank-order graph \hat{G} on V from the sequence of local rank-order graphs $\{G_k\}_{k=1}^{|S|}$. As explained in Sec. 3, the weights $\{p_k\}_{k=1}^{|S|}$ are set to one ($p_1 = \dots = p_{|S|} = 1$). In our experiments, the sampling set S is given by $S = \{k \in V : (k_y, k_x) \in S_x \times S_y\}$ where $S_x = \{1 + (i-1)ds\}_{i=1}^{n_1-ds}$ and $S_y = \{1 + (i-1)ds\}_{i=1}^{n_2-ds}$ define a downsampling grid with sampling step $ds = \lceil 2r(1-o) \rceil$, where the parameter o can take any value in the open interval $(0, 1)$.

4.2.2 Bilateral-filter graph

Instead of using rank-ordering information from \mathbf{Z} to construct the graph \hat{G} , we could use distance-ordering information. In doing so, we can adapt the graph construction procedure in Sec. 4.2.1 by allowing the sampling set $S = V$, and replacing step 3 with constructing a local distance-ordering graph $G_k = (V, E_k, w^{(k)})$, such that the edge set is given by $E_k = \{(k, j) : j \neq k, j \in \Omega_k\}$. Also, the weight function $w^{(k)} : V \times V \mapsto [0, 1]$ is defined by:

$$w^{(k)}(i, j) = \exp\left(-\frac{|i-j|^2}{2\delta_s^2}\right) \exp\left(-\frac{|Z_i - Z_j|^2}{2\delta_r^2}\right)$$

for $(i, j) \in E_k$, and $w^{(k)}(i, j) = 0$ otherwise, where δ_s and δ_r are the so-called bandwidth parameters of the bilateral filter. In our experiments, the bandwidth parameters are set as $\delta_s = 2r$, and $\delta_r = \text{mean}(\{\text{range}(Z_{\Omega_k})\}_{k=1}^{n_1 n_2})/6$, where $\text{range}(Z_{\Omega_k})$ computes the range of the pixel values of \mathbf{Z} at Ω_k .

5 NUMERICAL AND EXPERIMENTAL RESULTS

We now study the spectral-image reconstruction performance of the above mentioned approaches using simulated (ideal) and real data.

³We select Ω_k to be a square neighborhood, but it could be arbitrarily shaped.

5.1 Simulated Measurements

For simulated data, we selected a total of 25 hyperspectral image datacubes, including those in Fig. 9, captured at our lab, five datacubes from KAIST [31], eight datacubes from Harvard [32], and eight datacubes from ICVL [33], which are publically available databases⁴. For convenience, we downsampled the datacubes to the size $512 \times 512 \times L$, where L represents the number of frequency bands of a given datacube. We then generated a pair of snapshots of the datacubes, i.e., compressive and panchromatic images, using the forward models (33) and (34). The coded aperture of size 512×512 was set to have about 50% transmittance, with block and unblock pixels distributed uniformly at random. And the spectral sensitivities of the FPA1 and FPA2 were set to $s_1(l) = \{1\}_{l=1}^L$ and $s_2(l) = \{1\}_{l=1}^L$. Since the sensing matrix is perfectly known, any reconstruction errors will be mostly associated to the signal models encoding our prior knowledge of the signal of interest.

Given the snapshots and sensing matrices, the datacubes were reconstructed using the programs (35), (36), and (38), where the constrained optimization problems were solved using C-SALSA [34], and SYMMLQ [35], respectively. For each solver the noise level ϵ , the maximum number of iterations and stopping tolerance were set to zero, 10000, and 10^{-6} . The stopping criterion is satisfied if the relative change of the objective function is less than the tolerance.

The rank-order graph was constructed according to the procedure in Sec. 4.2.1 with radius and overlap parameters $r = 7$, and $o = \frac{3}{4}$. Likewise, the bilateral-filter graph was constructed with radius parameter $r = 3$ as indicated in Sec. 4.2.2. The weights in (37) were set as $w_1 = w_2 = \dots = w_L = 1$. Table 1 shows the average Peak Signal-to-Noise Ratio (PSNR), Spectral Angle Mapper (SAM), and the average computation time (CTIME) to reconstruct the scenes across different datacubes⁵.

In the table, it can be observed that all approaches outperform the WDCT approach, and the TV approach performs remarkably well in comparison with the proposed ROG and BFG approaches. However, our approach only uses rank-ordering information from the side information, and it doesn't use the side information as a direct measurement. In the discussion, we elaborate on why this may be advantageous in practice. If we were to do so, then we would improve the PSNR and SAM value of our reconstructions, but this would make the problem harder to solve by increasing the number of equations in the forward model.

As observed in Table 1, our approach is substantially faster than the other considered approaches because our method boils down to solving a sparse system of linear equations. In contrast, the TV and WDCT approaches use C-SALSA, which breaks down the associated constrained optimization programs into multiple easier programs, which may lead to delayed convergence [36].

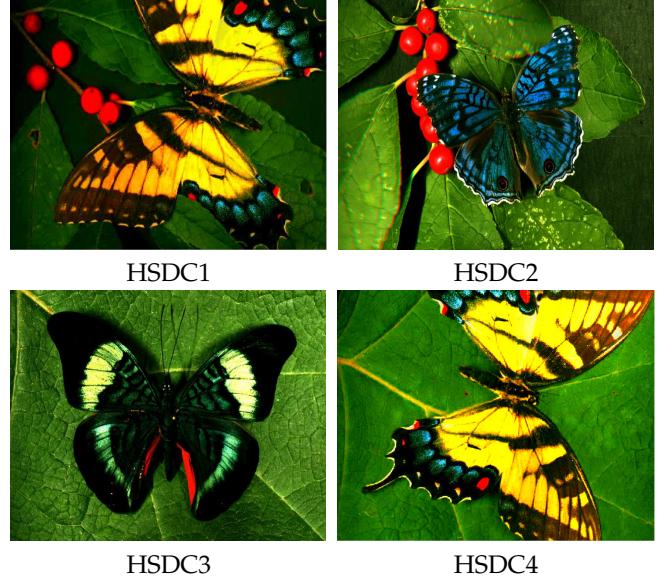


Fig. 9. UDEL hyperspectral image database. RGB renderings of four hyperspectral datacubes (HSDC) used for performance evaluation of the signal recovery algorithms. The hyperspectral datacubes HSDC1, HSDC2, HSDC3, and HSDC4 consist of 31 spectral bands of size 2064-by-3088 pixels, ranging from 400 to 700 nm. In the supplementary materials, we describe the experimental setup and methodology used to acquire these hyperspectral datacubes.

TABLE 1

Performance comparison between the reconstruction approaches ROG, BFG, TV, and WDCT across different databases UDEL, HRVRD, ICVL, KAIST. The approaches ROG and BFG refer to solutions found using the program (38), where the graph \hat{G} was constructed as described in Secs. (4.2.1) and (4.2.2), respectively. And the approaches TV and WDCT refer to solutions found using the programs (35) and (36).

		ROG	BFG	TV	WDCT
UDEL	PSNR (dB)	36.0± 2.0	36.0± 1.5	35.9± 1.4	29.7± 0.9
	SAM (deg)	9.9± 3.4	10.6± 4.0	10.0± 4.1	14.8± 3.9
	CTIME (min)	48.3± 9.6	173.3± 3.1	250.1± 8.9	232.1± 4.4
HRVRD	PSNR (dB)	40.1± 3.6	36.1± 2.5	39.3± 5.0	33.2± 3.7
	SAM (deg)	3.2± 1.0	3.3± 0.9	3.1± 1.1	7.4± 2.9
	CTIME (min)	60.6± 19.8	196.7± 35.7	250.2± 12.8	263.9± 38.1
ICVL	PSNR (dB)	36.9± 2.5	33.6± 0.7	37.6± 2.8	32.9± 3.2
	SAM (deg)	2.8± 1.2	2.7± 0.7	2.4± 0.7	4.7± 1.4
	CTIME (min)	49.6± 8.8	255.4± 59.8	266.2± 21.0	312.3± 10.4
KAIST	PSNR (dB)	36.4± 1.1	34.8± 1.1	42.0± 1.6	33.9± 2.0
	SAM (deg)	5.5± 1.1	4.0± 0.9	2.8± 0.7	7.6± 1.4
	CTIME (min)	51.1± 4.7	237.1± 15.7	230.7± 21.9	324.8± 11.4

5.2 Real Measurements

To assess our approach in a practical scenario, we implemented the DCCSI system in Fig. 8 as displayed in Fig. 10. Although our system resembles the dual-camera system proposed in [28], we note that our testbed provides a new experimental setting. Unlike the comparing system, the CASSI and side-information arms of our system contain non-identical focal plane arrays with different pixel-pitch sizes, which might make integration of the two imaging modalities slightly more challenging as will be discussed.

The coded aperture pixels are about eight times as big as the FPA1 pixels, so during alignment we set the relay lens to map the 128×128 coded-aperture pixels onto approximately

⁴We list the names of the selected datacubes in the supplementary materials

⁵The definitions of PSNR and SAM can be found in the Appendix D of the companion supplementary materials.

TABLE 2
Some specifications of the dual-camera compressive spectral imaging system in Fig. 10.

	Type	Resolution	Pitch size (μm)
FPA1	Basler ac A3088-56 μm	3088 × 2064	2.4 × 2.4
Coded aperture	Random, On-Off	128 × 128	19.6 × 19.6
FPA2	Allied Vision GC1290C	960 × 1280	3.75 × 3.75
Dispersive element	Double Amici prism	-	-

1024 × 1024 camera pixels. In addition to the specifications in Table 2, our system consists of an objective lens, CoastalOpt UV-VIS-IR 60 mm Macro, that images the spectral scene of interest onto a bandpass filter, which limits the spectral range of the scene to 450–650 nm. The filtered image is then relayed onto the CASSI's optical path and the side information camera by a 4-F system with a beam splitter located at the Fourier plane.

Under the above considerations, our DCCSI system matrix $\mathbf{H}_{\text{CASSI}}$ was configured to reconstruct 31 spectral bands of size 256-by-286 by forming super pixels of size 4 × 4. And the entries of $\mathbf{H}_{\text{CASSI}}$ were adjusted by following a slightly modified version of the calibration procedure proposed by [37]. The side-information camera needs to be co-registered with respect to the CASSI's camera and its spectral sensitivity has to be estimated. To do so, for each scene, we perform a preliminary reconstruction using total variation without side-information, and align the panchromatic image to the reconstructed panchromatic image using Matlab's image registration toolbox. The entries of \mathbf{R}_{PAN} (or equivalently its spectral sensitivity) can be estimated as in the CASSI's procedure. But unlike the CASSI's, for each calibration wavelength, each calibration image is collapsed into a single value by averaging all of the pixel values.

Figures 11 and 12 show the scenes of interest and their compressive and panchromatic snapshots as seen by our DCCSI system. Given the snapshots and the calibrated system matrices, reconstructions were performed using unconstrained formulations of the TV, BFG, and ROG approaches, presented in Sec. 5.1. **In the supplementary materials, we elaborate on such formulations.** The WDCT approach was

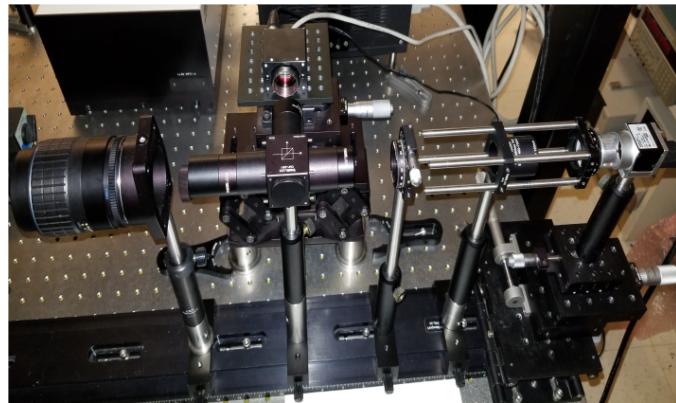


Fig. 10. Dual-camera compressive spectral imager.

excluded from the comparison since it exhibited inferior performance in our numerical experiments.

Similar to our previous experiment, the rank-order graph was constructed with radius and overlap parameters $r = 3$, and $o = \frac{3}{4}$. And the bilateral-filter graph was constructed with radius $r = 3$ and bandwidth parameters as indicated in Sec. 4.2.2. The weights in (37) were set as $w_l \propto 1/s_1(\lambda_l)$ to compensate for the non-uniform spectral response of the CASSI system.

The RGB renderings of the scenes SCN1, SCN2, SCN3, and SCN4 reconstructed by the TV, BFG, and ROG approaches can be observed in Fig. 13. The renderings were obtained using the standard color matching functions under the assumption of a flat illuminant power spectrum. Unlike the TV and BFG approaches, the proposed ROG approach produces RGB images, which not only contain as much spatial details as the associated panchromatic images but also display almost as many colors as the color images in Fig. 11. Note that the reconstructed RGB images are not meant to reproduce the exact same colors as the color images in Fig. 11. For high color fidelity, we must transform the reconstructed spectral datacubes into reflectance (or radiance) spectral datacubes, which involve an additional calibration process explained in [38]. In particular, one should remove the effects of the illuminant power spectrum and CASSI system's spectral response from the reconstructed datacubes.

The RGB renderings produced by our approach, however, exhibit chromatic artifacts similar to that of color bleeding, where color between nearby objects is transferred, which indicates imperfect decomposition of the spectral bands. Observe, for example, at the green leaves and red flowers of the SCN1, *blocks* dataset. The black background is contaminated by the green and red colors.

The chromatic artifacts must not only be attributed to the approximate rank-orderings induced by the panchromatic side image. In fact, we found that misalignment between the CASSI and side information cameras can also contribute to those artifacts. By using the quality index of the alignment procedure, explained in Sec. ??, we can rank-order the scenes from best to worst as follows: SCN2, SCN1, SNC4, SCN3. Note that chromatic artifacts appear to be less prevalent in the best registered scenes than in the worst ones. In addition to misalignment improvement, color-bleeding artifacts may be decreased by modifying the square neighborhoods (involved in the construction of the rank-order graphs) to the geometry of the objects in the scene. The bilateral-filter graph, for instance, appears to adapt better to the geometry, and therefore color bleeding seems less prominent.

Figure 14 displays the spectral datacube of the SCN2 as reconstructed by the TV, BFG, and ROG approaches. We observe that the proposed ROG approach leads to intensity variations across the spectrum that agree with the color of the original scene in addition to high spatial resolution. Although the BFG approach leads to similar performance, the reconstructed datacube misses to preserve some spatial details. On the other hand, the TV approach does not achieve the same spatial fidelity and exhibits undesirable behavior at the lower bands 440 nm, 442 nm, and 450 nm.

Figure 15 shows both reconstructed and reference spectral signatures at four different positions, P1, P2, P3, P4, in

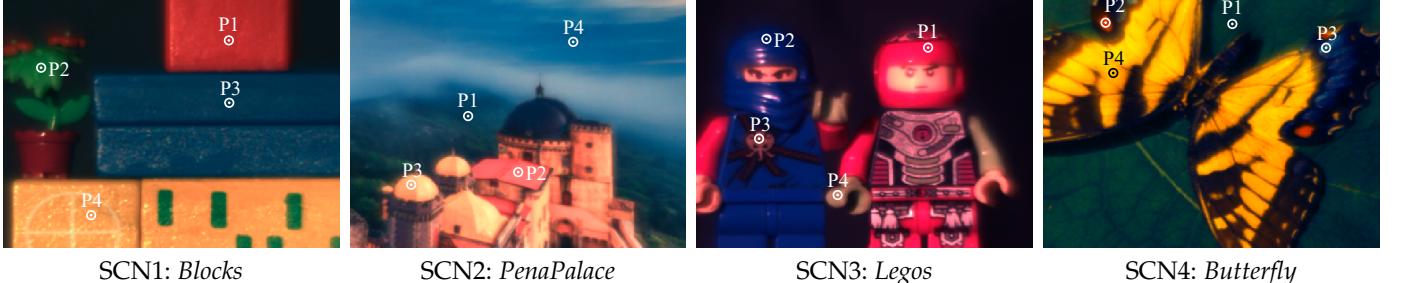


Fig. 11. Color images of the spectral scenes of interest. The color images were captured by the DCCSI system presented in Sec. 5.2, and subsequently processed to display more vivid colors. The spectral signatures at the points P1, P2, P3, and P4 as measured by a non-imaging spectrometer will be compared against reconstructed spectral signatures.

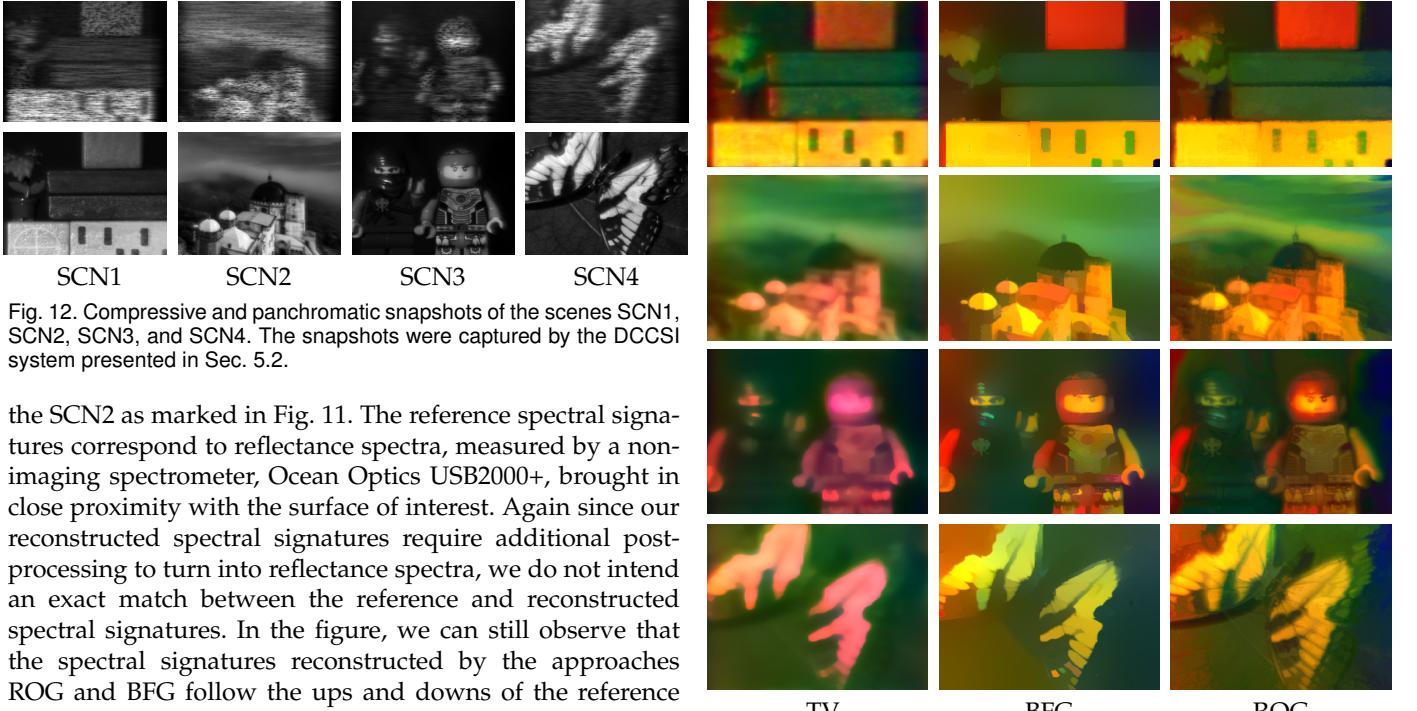


Fig. 12. Compressive and panchromatic snapshots of the scenes SCN1, SCN2, SCN3, and SCN4. The snapshots were captured by the DCCSI system presented in Sec. 5.2.

the SCN2 as marked in Fig. 11. The reference spectral signatures correspond to reflectance spectra, measured by a non-imaging spectrometer, Ocean Optics USB2000+, brought in close proximity with the surface of interest. Again since our reconstructed spectral signatures require additional post-processing to turn into reflectance spectra, we do not intend an exact match between the reference and reconstructed spectral signatures. In the figure, we can still observe that the spectral signatures reconstructed by the approaches ROG and BFG follow the ups and downs of the reference spectra, and outperform the TV approach in terms of the Root Mean Square Error (RMSE).

In the supplementary materials, we include the reconstructed spectral datacubes as well as comparisons between reconstructed and reference spectral signatures for the remaining scenes SCN1, SCN3, and SCN4.

6 DISCUSSION

In this paper, we develop the concept of rank-order graph and address the problem of signal recovery from undersampled linear measurements by using smoothness with respect to rank-order graphs. In particular, we show the usefulness of rank-order graphs for compressive spectral imaging, and demonstrate its advantages over some traditional approaches.

Despite having a relatively simple sparse edge structure, rank-order graphs are capable of capturing structural details in the data, leading to efficient graph representations. The key idea is that by rearranging the original feature space, the complexity of the rearranged space can be captured by using fundamental graphs, i.e., path graphs. In image processing applications such as denosing, super-resolution, inpainting and compression, a similar concept has shown promising results [13, 39].

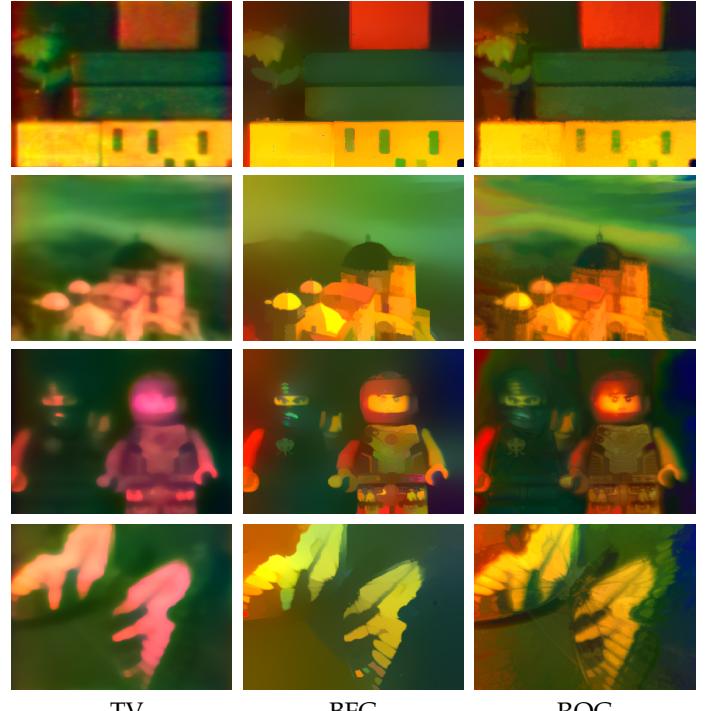


Fig. 13. Reconstructed raw RGB renderings of the scenes in Fig. 11. The RGB renderings, from left to right, are color representations of the reconstructed spectral datacubes obtained by the TV, BFG, and ROG approaches.

The fact that a one-dimensional projection, e.g. panchromatic image, from a high-dimensional object, e.g. hyperspectral image, can be used to rank order its elements smoothly appears to be rooted in the theory of manifolds. In our application scenario, we could say that if the hyperspectral image lives close to a low dimensional linear manifold, then the ordering information from the panchromatic image is useful to infer the order of the high dimensional vectors (or spectral signatures). We are currently exploring this hypothesis.

One of the simplest ways to integrate the side information in the CSI reconstruction process is as a direct measurement as in the TV approach. However, in our practical scenario, this approach failed to transfer the spatial details in the PAN image to the reconstructed spectral scene. Apparently, since the compressive and panchromatic images are captured by different detector arrays, with distinct spectral sensitivities, the calibrated forward model rendered too inconsistent and unable to force the reconstructed datacube to map closely onto the compressive and panchromatic

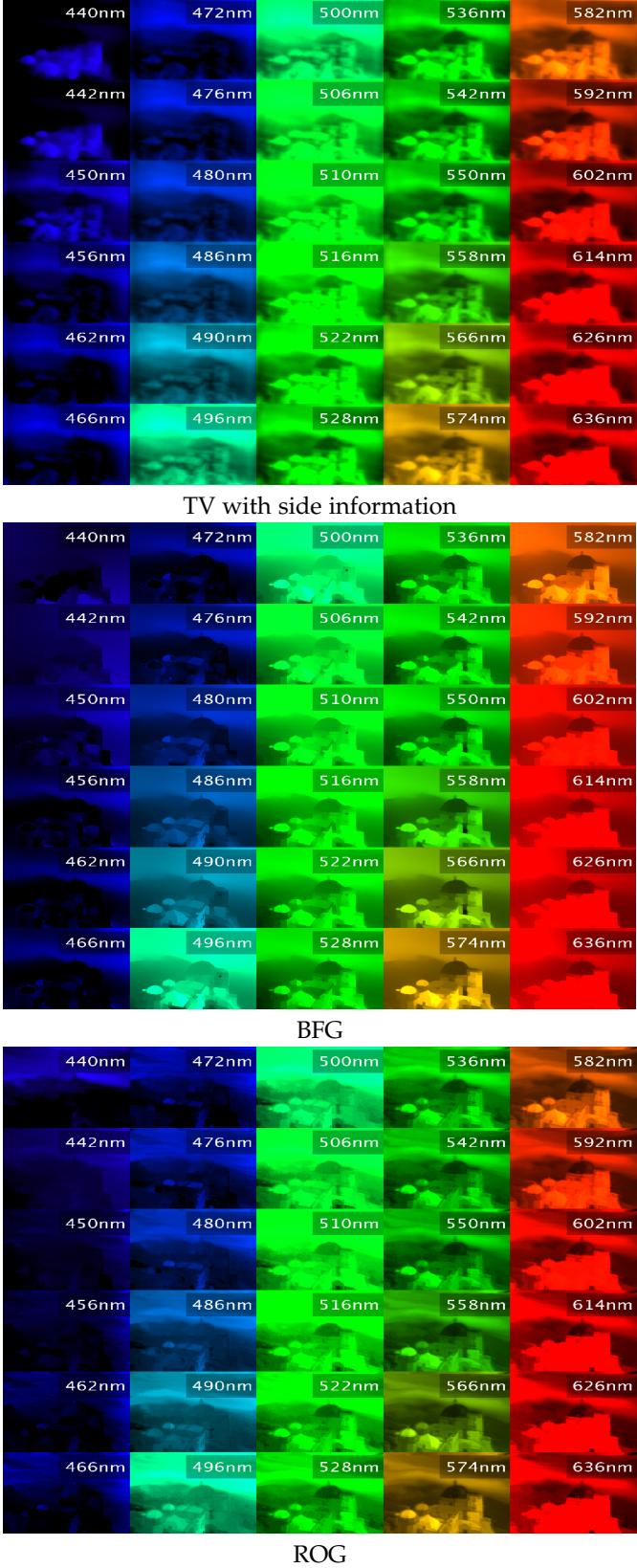


Fig. 14. Reconstructed spectral images of the scene SCN2 in Fig. 11. We display 30 out of 31 spectral bands of size 256×286 , obtained from top to bottom by the TV, BFG, and ROG approaches.

snapshots simultaneously. Because the panchromatic image provides merely rank-ordering information and does not play the role of direct measurement, our approach goes around such complications remarkably well.

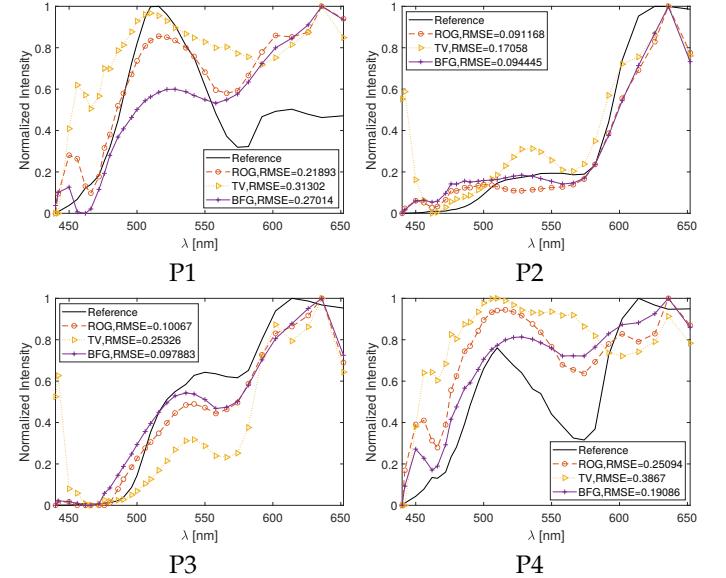


Fig. 15. Reconstructed and reference spectral signatures of SCN2 at the points P1, P2, P3, and P4, indicated in Fig. 11. The reconstructed spectra were approximated by taking the median value of a 5-by-5 window around points P1, P2, P3, P4 at each spectral band. The reference spectrum was obtained by a non-imaging spectrometer. Please, before judging the quality of the reconstructions by their apparent difference, consider reading the figure's explanation.

Rank-order graphs could be extended to applications such as depth map restoration [40], positron emission tomography enhancement [41], and spectral X-Ray computer tomography reconstruction [42] to name a few, where side information is readily available. To construct the graphs, however, we should not use rank ordering information, directly extracted from the side information itself. Instead, rank orderings have to consider the geometrical aspects of the multimodal feature space. In [43], the authors addressed the problem of ranking of multivariate vectors, which may be useful to generalize the notion of rank-order graphs. We are currently investigating these new application scenarios.

Adaptive compressive spectral imaging architectures may benefit from the use of rank-order graphs because not only does its construction take little computational overhead but their edge structure is also highly sparse. As suggested by Theorem 2, the approximation error is inversely related to the restricted strong convexity constant, which can be shown proportional to the smallest eigenvalue of the matrix $\mathbf{A}^T \mathbf{A} + \alpha \mathbf{L}_G^T \mathbf{L}_G$, where $\alpha > 0$ is a regularization constant. By allowing the coded aperture to modify the distribution of zeros and ones in the matrix \mathbf{A} such that the smallest eigenvalue is maximized, we may obtain richer spatio-spectrally coded measurements and therefore better reconstructions than random coded apertures. To explore this idea, one could follow a similar approach to that in [44].

We conclude by saying that rank-order graphs have the potential to become a valuable asset not only in modern-day computer vision and computational imaging problems, but also as models to learn efficient graph representations of data.

APPENDIX A

Proof of proposition 1. We have $\mathbf{L}_G = \Psi \Lambda \Psi^T$ with $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. So $\|\mathbf{L}_G \mathbf{x}\|_2 = \|\Psi \Lambda \Psi^T \mathbf{x}\|_2 = \|\Lambda \Psi^T \mathbf{x}\|_2$,

where $\|\Lambda \Psi^T \mathbf{x}\|_2 = (\sum_{i=1}^n \lambda_i^2 \langle \psi_i, \mathbf{x} \rangle^2)^{1/2}$. Using Jensen's inequality:

$$\begin{aligned} \|\mathbf{L}_G \mathbf{x}\|_2 &= (\sum_{i=1}^n \lambda_i^2)^{1/2} \left(\sum_{i=1}^n \frac{\lambda_i^2}{\sum_{i=1}^n \lambda_i^2} \langle \psi_i, \mathbf{x} \rangle^2 \right)^{1/2} \\ &\geq (\sum_{i=1}^n \lambda_i^2)^{1/2} \sum_{i=1}^n \frac{\lambda_i^2}{\sum_{i=1}^n \lambda_i^2} (\langle \psi_i, \mathbf{x} \rangle^2)^{1/2} \\ &= \sum_{i=1}^n \frac{\lambda_i^2}{(\sum_{i=1}^n \lambda_i^2)^{1/2}} |\langle \psi_i, \mathbf{x} \rangle| = \|\mathbf{W} \Psi^T \mathbf{x}\|_1. \end{aligned}$$

APPENDIX B

Proof of theorem 1. Forward proof. For the sake of contradiction, assume $\mathcal{N}_A \cap \mathcal{M}_{2\varepsilon} \neq \{\mathbf{0}\}$. Let $\mathbf{0} \neq \mathbf{z} \in \mathcal{N}_A \cap \mathcal{M}_{2\varepsilon}$. Define $\mathbf{x}_1 = -\mathbf{z}/2$ and $\mathbf{x}_2 = \mathbf{z}/2$. Clearly $\mathbf{x}_1 \neq \mathbf{x}_2$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{N}_A \cap \mathcal{M}_\varepsilon$. Since \mathbf{x}_1 and \mathbf{x}_2 belong in \mathcal{N}_A , we have that $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_2$. Conversely, if $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_2$ with $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{M}_\varepsilon$, then $\mathbf{x}_1 - \mathbf{x}_2 \in \mathcal{N}_A \cap \mathcal{M}_{2\varepsilon}$, and so $\mathbf{x}_1 = \mathbf{x}_2$.

APPENDIX C

Proof of lemma 1. Let $\varepsilon > 0$. For the sake of contradiction, assume $\forall \gamma > 0$, there exist a nonzero vector $\mathbf{v} \in \mathcal{N}_A$ such that $\gamma > \frac{\|\mathbf{L}_G \mathbf{v}\|_2}{\|\mathbf{v}\|_2}$. Set $\gamma := \varepsilon$. Then there exist a nonzero vector $\mathbf{x} \in \mathcal{N}_A$ such that $\frac{\|\mathbf{L}_G \mathbf{x}\|_2}{\|\mathbf{x}\|_2} < \gamma = \varepsilon$. Note that $\tilde{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|_2$ produces $\|\mathbf{L}_G \tilde{\mathbf{x}}\|_2 < \varepsilon$, thus by definition $\tilde{\mathbf{x}} \in \mathcal{M}_{2\varepsilon}$. But, this contradicts the hypothesis that $\mathcal{N}_A \cap \mathcal{M}_{2\varepsilon} = \{\mathbf{0}\}$. Therefore, it has to be true that there exist $\gamma > 0$, which is a lower bound of (8) for all nonzero $\mathbf{v} \in \mathcal{N}_A$.

APPENDIX D

Proof of theorem 2. Let $\mathbf{v} = \hat{\mathbf{x}} - \mathbf{x}$. By optimality of $\hat{\mathbf{x}}$, $\|\mathbf{L}_G(\mathbf{x} + \mathbf{v})\|_2 \leq \|\mathbf{L}_G \mathbf{x}\|_2$. This implies that $\|\mathbf{L}_G \mathbf{v}\|_2^2 + 2\langle \mathbf{L}\mathbf{x}, \mathbf{L}\mathbf{v} \rangle \leq 0$. Since $\|\mathbf{L}_G \mathbf{v}\|_2^2 \geq 0$, it must be that $\langle \mathbf{L}_G \mathbf{x}, \mathbf{L}_G \mathbf{v} \rangle = -|\langle \mathbf{L}_G \mathbf{x}_0, \mathbf{L}_G \mathbf{v} \rangle|$. By applying the Cauchy-Schwartz inequality, we obtain $\|\mathbf{L}_G \mathbf{v}\|_2^2 - 2\|\mathbf{L}_G \mathbf{x}_0\|_2 \|\mathbf{L}_G \mathbf{v}\|_2 \leq 0$. Now, by factoring out $\|\mathbf{L}_G \mathbf{v}\|_2$, the l.h.s term of the inequality becomes $\|\mathbf{L}_G \mathbf{v}\|_2 (\|\mathbf{L}_G \mathbf{v}\|_2 - 2\|\mathbf{L}_G \mathbf{x}_0\|_2)$, which implies $\|\mathbf{L}_G \mathbf{v}\|_2 - 2\|\mathbf{L}_G \mathbf{x}_0\|_2 \leq 0$. Using Lemma 1, we have $\gamma \|\mathbf{v}\|_2 \leq \|\mathbf{L}_G \mathbf{v}\|_2$. Then $\gamma \|\mathbf{v}\|_2 - 2\|\mathbf{L}_G \mathbf{x}_0\|_2 \leq 0$. Therefore we find that $\|\mathbf{v}\|_2 \leq 2\|\mathbf{L}_G \mathbf{x}_0\|_2 / \gamma$, which concludes the proof.

APPENDIX E

The proof of theorem 3 depends on the following lemma.

Lemma 2. Let $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$. Then

$$\min_{\sigma \in S_n} \sum_{i=1}^{n-1} (x_{\sigma_{i+1}} - x_{\sigma_i})^2 = \sum_{i=1}^{n-1} (x_{\hat{\sigma}_{i+1}} - x_{\hat{\sigma}_i})^2,$$

where $\hat{\sigma} \in S_n$ is such that $x_{\hat{\sigma}_1} \leq x_{\hat{\sigma}_2} \leq \dots \leq x_{\hat{\sigma}_n}$.

Proof of lemma 2. Replacing \mathbf{x} by $\mathbf{x}_{\hat{\sigma}} = (x_{\hat{\sigma}_1}, \dots, x_{\hat{\sigma}_n})^T$, we may assume without loss of generality that $x_1 \leq x_2 \leq \dots \leq x_n$. Let $\sigma \in S_n$, and $1 \leq i \leq n-1$. Let $t_{i+1} := \max(\sigma_{i+1}, \sigma_i)$. Then $(x_{\sigma_{i+1}} - x_{\sigma_i})^2 \geq (x_{t_{i+1}} - x_{t_{i+1}-1})^2$. Summing over i , we conclude that $\sum_{i=1}^{n-1} (x_{\sigma_{i+1}} - x_{\sigma_i})^2 \geq \sum_{i=1}^{n-1} (x_{t_{i+1}} - x_i)^2$.

Proof of theorem 3. Recall that for any graph $G = (V, E)$ and any vector $\mathbf{z} = (z_1, \dots, z_n)^T \in \mathbb{R}^n$, we have

$$\mathbf{z}^T \mathbf{L}_G \mathbf{z} = \sum_{(i,j) \in E(G)} (z_i - z_j)^2.$$

Thus, for any path $(i_1, i_2), (i_2, i_3), \dots, (i_{n-1}, i_n)$ on V , we have $\mathbf{x}^T \mathbf{L}_G \mathbf{x} = \sum_{(i_k, i_{k+1}) \in E(G)} (x_{i_k} - x_{i_{k+1}})^2$.

The set \mathcal{L}_n is in bijection with permutations in S_n by identifying the path with edges $(i_1, i_2), (i_2, i_3), \dots, (i_{n-1}, i_n)$ with the permutation $\sigma \in S_n$ where $\sigma_k = i_k$. Thus,

$$\min_{\sigma \in \mathcal{L}_n} \mathbf{x}^T \mathbf{L}_G \mathbf{x} = \min_{\sigma \in S_n} \sum_{i=1}^{n-1} (x_{\sigma_{i+1}} - x_{\sigma_i})^2$$

The result now follows from Lemma 2.

APPENDIX F

Proof of theorem 4. By the triangle inequality, we have

$$\|\mathbf{L}_G \mathbf{x} - \mathbf{L}_{G^*} \mathbf{x}\|_2 \leq \|\mathbf{L}_G \mathbf{x}\|_2 + \|\mathbf{L}_{G^*} \mathbf{x}\|_2.$$

Recall that for any graph G , its graph Laplacian \mathbf{L}_G can be written as $\mathbf{L}_G = \mathbf{M}_G^T \mathbf{M}_G$, where \mathbf{M}_G^T is the incidence matrix of G . By definition of matrix norm, we have $\|\mathbf{L}_G \mathbf{x}\|_2 \leq \|\mathbf{M}_G^T\|_2 (\mathbf{x}^T \mathbf{L}_G \mathbf{x})^{1/2}$. Since $G, G^* \in \mathcal{L}_n$, $\|\mathbf{M}_G^T\|_2$ is equal to $\|\mathbf{M}_{G^*}^T\|_2$. We obtain

$$\|\mathbf{L}_G \mathbf{x} - \mathbf{L}_{G^*} \mathbf{x}\|_2 \leq 2\|\mathbf{M}_G^T\|_2 ((\mathbf{x}^T \mathbf{L}_G \mathbf{x})^{1/2} + (\mathbf{x}^T \mathbf{L}_{G^*} \mathbf{x})^{1/2}).$$

From Theorem 3, we have $\mathbf{x}^T \mathbf{L}_{G^*} \mathbf{x} \leq \mathbf{x}^T \mathbf{L}_G \mathbf{x}$. Thus,

$$\|\mathbf{L}_G \mathbf{x} - \mathbf{L}_{G^*} \mathbf{x}\|_2 \leq 4\|\mathbf{M}_G^T\|_2 (\mathbf{x}^T \mathbf{L}_G \mathbf{x})^{1/2}.$$

The result follows from proving that $\|\mathbf{M}_G^T\|_2 \leq 2$ and applying the triangle inequality to $(\mathbf{x}^T \mathbf{L}_G \mathbf{x})^{1/2}$.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants NSF 1815992 and NSF 1816003. Juan F. Florez thanks Fulbright Colombia and Colciencias for his doctoral fellowship, and Hoover Rueda and Alejandro Parada for insightful technical discussions.

REFERENCES

- [1] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge university press, 2012.
- [2] L. Sun, Z. Fan, X. Fu, Y. Huang, X. Ding, and J. Paisley, "A deep information sharing network for multi-contrast compressed sensing mri reconstruction," *IEEE Trans. on Image Processing*, vol. 28, no. 12, pp. 6141–6153, 2019.
- [3] N. Allouche, K. Øzdemir, A. Özbek, and J.-F. Hopperstad, "Multimeasurement aliased-noise attenuation for sparse land seismic acquisition using compressed sensing," *Geophysics*, vol. 85, no. 2, pp. V183–V200, 2020.
- [4] Y. Á. López and J. Á. M. Lorenzo, "Compressed sensing techniques applied to ultrasonic imaging of cargo containers," *Sensors*, vol. 17, no. 1, p. 162, 2017.

- [5] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, NY, USA; 2001, vol. 1, no. 10.
- [6] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.
- [7] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948.
- [8] J. Yang, X. Liao, X. Yuan, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Compressive sensing by learning a gaussian mixture model from measurements," *IEEE Trans. on Image Processing*, vol. 24, no. 1, pp. 106–119, 2014.
- [9] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *Proceedings of the 34th International Conference on Machine Learning–Volume 70*. JMLR.org, 2017, pp. 537–546.
- [10] D. Gilton, G. Ongie, and R. Willett, "Neumann networks for linear inverse problems in imaging," *IEEE Trans. on Computational Imaging*, 2019.
- [11] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.
- [12] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *arXiv preprint arXiv:1912.10557*, 2019.
- [13] G. Vaksman, M. Zibulevsky, and M. Elad, "Patch ordering as a regularization for inverse problems in image processing," *SIAM Journal on Imaging Sciences*, vol. 9, no. 1, pp. 287–319, 2016.
- [14] F. Olken and D. Rotem, "Rearranging data to maximize the efficiency of compression," *Journal of Computer and System Sciences*, vol. 38, no. 2, pp. 405–430, 1989.
- [15] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 105–115, 2014.
- [16] M.-Y. Chen, F. Renna, and M. R. Rodrigues, "Compressive sensing with side information: How to optimally capture this extra information for gmm signals?" *IEEE Trans. on Signal Processing*, vol. 66, no. 9, pp. 2314–2329, 2018.
- [17] X. Yuan, T.-H. Tsai, R. Zhu, P. Llull, D. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," *IEEE Journal of selected topics in Signal Processing*, vol. 9, no. 6, pp. 964–976, 2015.
- [18] L. Galvis, D. Lau, X. Ma, H. Arguello, and G. R. Arce, "Coded aperture design in compressive spectral imaging based on side information," *Applied optics*, vol. 56, no. 22, pp. 6332–6340, 2017.
- [19] L. Wang, Z. Xiong, H. Huang, G. Shi, F. Wu, and W. Zeng, "High-speed hyperspectral video acquisition by combining nyquist and compressive sampling," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 41, no. 4, pp. 857–870, 2019.
- [20] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [21] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.
- [22] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.
- [23] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 44–63, 2019.
- [24] H. Dym, *Linear algebra in action*. American Mathematical Soc., 2013, vol. 78.
- [25] G. Strang, "The discrete cosine transform," *SIAM review*, vol. 41, no. 1, pp. 135–147, 1999.
- [26] K.-S. Lu and A. Ortega, "Symmetric line graph transforms for inter predictive video coding," in *2016 Picture Coding Symposium (PCS)*. IEEE, 2016, pp. 1–5.
- [27] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied optics*, vol. 47, no. 10, pp. B44–B51, 2008.
- [28] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," *Applied optics*, vol. 54, no. 4, pp. 848–858, 2015.
- [29] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 2539–2543.
- [30] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse problems and imaging*, vol. 2, no. 4, pp. 455–484, 2008.
- [31] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8032–8041.
- [32] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *CVPR 2011*. IEEE, 2011, pp. 193–200.
- [33] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *European Conference on Computer Vision*. Springer, 2016, pp. 19–34.
- [34] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. on Image Processing*, vol. 20, no. 3, pp. 681–695, 2010.
- [35] C. C. Paige and M. A. Saunders, "Solution of sparse indefinite systems of linear equations," *SIAM journal on numerical analysis*, vol. 12, no. 4, pp. 617–629, 1975.
- [36] J. Zhang, Y. Peng, W. Ouyang, and B. Deng, "Accelerating admm for efficient simulation and optimization," *ACM Trans. on Graphics (TOG)*, vol. 38, no. 6, pp. 1–21, 2019.
- [37] A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Spectral image estimation for coded aperture

- snapshot spectral imagers," in *Image Reconstruction from Incomplete Data V*, vol. 7076. International Society for Optics and Photonics, 2008, p. 707602.
- [38] D. H. Foster and K. Amano, "Hyperspectral imaging in color vision research: tutorial," *JOSA A*, vol. 36, no. 4, pp. 606–627, 2019.
- [39] R. Budinich, "A region-based easy-path wavelet transform for sparse image representation," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 15, no. 05, p. 1750045, 2017.
- [40] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, and W. Gao, "Depth restoration from rgb-d data via joint adaptive regularization and thresholding on manifolds," *IEEE Trans. on Image Processing*, vol. 28, no. 3, pp. 1068–1079, 2018.
- [41] M. J. Ehrhardt, P. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J. S. Duncan, L. Pizarro, D. Atkinson, B. F. Hutton, S. Ourselin *et al.*, "Pet reconstruction with an anatomical mri prior using parallel level sets," *IEEE Trans. on medical imaging*, vol. 35, no. 9, pp. 2189–2199, 2016.
- [42] Q. Yang, W. Cong, Y. Xi, and G. Wang, "Spectral x-ray ct image reconstruction with a combination of energy-integrating and photon-counting detectors," *PloS one*, vol. 11, no. 5, p. e0155374, 2016.
- [43] R. C. Hardie and G. R. Arce, "Ranking in \mathbb{R}^p and its use in multivariate image estimation," *IEEE Trans. on circuits and systems for video technology*, vol. 1, no. 2, pp. 197–209, 1991.
- [44] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, and W. Gao, "Fast graph sampling set selection using gershgorin disc alignment," *IEEE Trans. on Signal Processing*, vol. 68, pp. 2419–2434, 2020.