A MULTIPLE MEASUREMENT VECTOR APPROACH TO SYNTHETIC APERTURE RADAR IMAGING

LILIANA BORCEA AND ILKER KOCYIGIT *

Abstract. We study a multiple measurement vector (MMV) approach to synthetic aperture radar (SAR) imaging of scenes with direction dependent reflectivity and with polarization diverse measurements. The data are gathered by a moving transmit-receive platform which probes the imaging scene with signals and records the backscattered waves. The unknown reflectivity is represented by a matrix with row support corresponding to the location of the scatterers in the scene, and columns corresponding to measurements gathered from different sub-apertures, or different polarization of the waves. The MMV methodology is used to estimate the reflectivity matrix by inverting in an appropriate sense the linear system of equations that models the SAR data. We obtain a resolution analysis of SAR imaging with MMV, which takes into account the sparsity of the imaging scene, the separation of the scatterers and the diversity of the measurements. The results of the analysis are illustrated with some numerical simulations.

Key words. synthetic aperture radar imaging, convex optimization, multiple measurement vector, simultaneously sparse.

AMS subject classifications. 35Q93, 58J90, 45Q05.

1. Introduction. Sparsity promoting optimization [26, 25, 22, 9, 11, 10, 12] is an important methodology for imaging applications where scenes that are sparse in some representation can be reconstructed with high resolution. There is a large body of literature on this topic in synthetic aperture radar imaging [4, 32, 28], sensor array imaging [13, 14, 7], medical imaging [30], astronomy [6], geophysics [33], and so on.

We are interested in the application of synthetic aperture radar (SAR) imaging, where a transmit-receive antenna mounted on a moving platform probes an imaging scene with waves and records the scattered returns [20, 17]. This is a particular inverse problem for the wave equation, where the waves propagate through a homogeneous medium, back and forth between the platform and the imaging scene, and the unknown is modeled as a two-dimensional reflectivity function of location on a known imaging surface. Most SAR imaging is based on a linear model of the data, where the unknown reflectivity is represented by a collection of independent point scatterers [17]. The image is then formed by inverting approximately this linear relation, using filtered backprojection or matched filtering [17], also known as Kirchhoff migration [5]. Such imaging is popular because it is robust to noise, it is simple and works well when the linear model is a good approximation of the data. However, the resolution is limited by the extent of the aperture, the frequency and the bandwidth of the probing signals emitted by the moving platform [20, 17]. The promise of sparsity promoting optimization is that these resolution limits can be overcome when the unknown reflectivity has sparse support [4, 32, 28].

The modeling of the reflectivity as a collection of points that scatter the waves isotropically may lead to image artifacts. It is known that even if the scatterers are small, so that their support may be represented by a point and the single scattering approximation (i.e., the linear data model) can be used, their reflectivity may depend on the frequency and the direction of illumination [2, Chapters 3, 5]. Moreover, the scatterers have an effective polarization tensor that describes their response to different polarizations of the probing electromagnetic waves [2, 3]. Thus, the reflectivity function depends on more variables than the two dimensional location vector assumed in conventional SAR, and depending on how strong this dependence is, the resulting images may be worse than expected. For example, a scatterer that reflects only within a narrow cone of incident angles cannot be sensed over most of the synthetic aperture, so its reconstruction with filtered backprojection will have low resolution. Direct application of sparse optimization methods does not give good results either, because of the large systematic error in the linear data model that assumes a scalar, constant reflectivity over the entire aperture.

SAR imaging of frequency-dependent reflectivities has been studied in [16, 35, 34], using either Doppler effects, or data segmentation over frequency sub-bands. Data segmentation is a natural idea for imaging both frequency and direction dependent reflectivities that are regular enough so that they can be approximated as piecewise constant functions over properly chosen frequency sub-bands and cones of angles of incidence (i.e.,

---

*Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043.
Email: borcea@umich.edu & ilkerk@umich.edu

sub-apertures). Images can be obtained separately from each data segment, but the question is how to fuse the information to achieve better resolution. The study in [8] uses the multiple measurement vector (MMV) methodology [31, 15, 39], also known as simultaneously sparse approximation [38, 37], for this purpose. The MMV framework fits here because the reflectivity is supported at the same locations in the imaging scene, for each data set. Only the values of the reflectivity change. In the discrete setting, this means that the unknown is represented by a matrix $\boldsymbol{X}$ with row support corresponding to the pixels in the image that contain scatterers, and with columns corresponding to the different values of the reflectivity for each frequency band and sub-aperture. The MMV methodology is used in [8] to reconstruct this unknown matrix under the assumption of sparse row support.

In this paper we pursue the ideas in [8] further, by studying the resolution of the MMV reconstructions and analyzing conditions under which the multiple views of the imaging scene improve the results. We also discuss the application of the MMV methodology to SAR imaging with polarization diverse measurements.

The paper is organized as follows: We begin in section 2 with the theoretical results, stated for a general linear system with unknown matrix $\boldsymbol{X}$. They consist of a resolution theory of imaging with MMV, that takes into account the separation of the points in the support of the reflectivity. We also show that if the rows of $\boldsymbol{X}$ are almost orthogonal, we can expect better reconstructions than with sparsity promoting ℓ_1 optimization applied separately to each data set. This orthogonality condition may arise in SAR imaging, as explained in section 3, in the context of imaging direction dependent reflectivities. SAR with polarization diverse measurements is discussed in section 4. The proofs of the results are in section 5. We end with a summary in section 6.

2. Theory. We state here our main results on the resolution of imaging with MMV. We begin in section 2.1 with a brief discussion on MMV, and then give the results in section 2.2.

We use throughout the following notation convention: Bold uppercase letters, as in $\boldsymbol{X} \in \mathbb{C}^{N_{\boldsymbol{y}} \times N_v}$ denote matrices, and bold lowercase letters denote vectors. We also use an arrow index, as in $\boldsymbol{x}_{j\rightarrow} \in \mathbb{C}^{1 \times N_v}$, to distinguish the rows of $\boldsymbol{X}$ from its column vectors denoted by $\boldsymbol{x}_j \in \mathbb{C}^{N_{\boldsymbol{y}} \times 1}$.

2.1. Preliminaries. Consider a general linear model of a data matrix $\boldsymbol{D} \in \mathbb{C}^{N_{\boldsymbol{r}} \times N_v}$,

$$\boldsymbol{GX} = \boldsymbol{D}, \tag{2.1}$$

where the unknown matrix $\boldsymbol{X} \in \mathbb{C}^{N_{\boldsymbol{y}} \times N_v}$ is mapped to $\boldsymbol{D}$ by a given sensing matrix $\boldsymbol{G} \in \mathbb{C}^{N_{\boldsymbol{r}} \times N_{\boldsymbol{y}}}$. In the context of SAR imaging, $\boldsymbol{X}$ represents the unknown reflectivity discretized* at $N_{\boldsymbol{y}}$ points $\{\boldsymbol{y}_j\}_{1 \leq j \leq N_{\boldsymbol{y}}}$ in the imaging region Ω, a bounded set on a known surface. The matrix $\boldsymbol{D}$ is an aggregate of N_v data sets or views, each consisting of $N_{\boldsymbol{r}}$ measurements of the wave field at the moving radar antenna. A column of $\boldsymbol{X}$, denoted generically by $\boldsymbol{x}_v$ for $1 \leq v \leq N_v$, represents the reflectivity for a given view, and it is mapped to the corresponding column $\boldsymbol{d}_v$ of $\boldsymbol{D}$ by the sensing matrix $\boldsymbol{G}$. This $\boldsymbol{G}$ is the discretization of the kernel of the linear integral operator that defines the single scattering (Born) approximation of the solution of the wave equation, as described in section 3. Its entries are given by Green's functions which model the wave propagation between the imaging points in Ω and the locations of the antenna in the synthetic aperture.

When $N_v = 1$, the linear model (2.1) corresponds to the single measurement vector (SMV) problem,

$$\boldsymbol{Gx} = \boldsymbol{d}, \tag{2.2}$$

with unknown vector $\boldsymbol{x} \in \mathbb{C}^{N_{\boldsymbol{y}} \times 1}$ and data vector $\boldsymbol{d} \in \mathbb{C}^{N_{\boldsymbol{r}} \times 1}$, where we dropped the column index 1. The SMV problem has been studied extensively in the context of compressed sensing [26, 25, 22, 9, 11, 10, 12, 29] for the undetermined case $N_{\boldsymbol{r}} \ll N_{\boldsymbol{y}}$. In particular, it is known [23, Corollary 1] that if[†]

$$\|\boldsymbol{x}\|_0 < \frac{\text{spark}(\boldsymbol{G})}{2}, \tag{2.3}$$

where $\|\boldsymbol{x}\|_0$ equals the number of nonzero entries in $\boldsymbol{x}$, then (2.2) has a unique solution satisfying (2.3), given by the minimizer of the combinatorial optimization problem

$$\text{minimize } \|\boldsymbol{z}\|_0 \text{ subject to } \boldsymbol{Gz} = \boldsymbol{d}. \tag{2.4}$$

---

*We assume that the $N_{\boldsymbol{y}}$ points define a fine mesh in Ω, so we can neglect errors due to scatterer locations off the mesh.
[†]Recall from [23] that spark($\boldsymbol{G}$) is the smallest number of linearly dependent columns of $\boldsymbol{G}$.

This result is generalized in [15, Theorem 2.4] to the MMV problem (2.1) for $N_v > 1$. It states that when the number of nonzero rows in $\boldsymbol{X}$, denoted by $\|\boldsymbol{X}\|_0$, satisfies

$$\|\boldsymbol{X}\|_0 < \frac{\operatorname{spark}(\boldsymbol{G}) + \operatorname{rank}(\boldsymbol{D}) - 1}{2}, \tag{2.5}$$

the linear system (2.1) has a unique solution satisfying (2.5), given by the minimizer of the combinatorial optimization problem

$$\text{minimize } \|\boldsymbol{Z}\|_0 \text{ subject to } \boldsymbol{G}\boldsymbol{Z} = \boldsymbol{D}. \tag{2.6}$$

Thus, if the different data sets bring new information, so that $\boldsymbol{D}$ has large rank, the MMV problem is uniquely solvable for less stringent conditions on the row support of $\boldsymbol{X}$ i.e., for less sparse imaging scenes.

The combinatorial problems (2.4) and (2.6) are not computationally tractable, so they are replaced by convex relaxations. The minimizer of the convex problem

$$\mathscr{P}_1: \text{ minimize } \|\boldsymbol{z}\|_1 \text{ subject to } \boldsymbol{G}\boldsymbol{z} = \boldsymbol{d}, \tag{2.7}$$

where $\|\cdot\|_1$ is the ℓ_1 norm, is known to give the exact solution $\boldsymbol{x}$ of the SMV problem (2.2) under various conditions satisfied by $\boldsymbol{G}$ and $\boldsymbol{x}$, like the null space property [18], the restricted isometry property [10], conditions based on the mutual coherence [24] and the cumulative coherence [36]. Relaxations of (2.6) of the form

$$\mathscr{P}_{1,q}: \text{ minimize } \|\boldsymbol{Z}\|_{1,q} \text{ subject to } \boldsymbol{G}\boldsymbol{Z} = \boldsymbol{D}, \tag{2.8}$$

are studied in [19, 31, 15, 27, 38, 37, 39] using the $\ell_{1,q}$ norm

$$\|\boldsymbol{Z}\|_{1,q} = \sum_{j=1}^{N_{\boldsymbol{y}}} \|\boldsymbol{z}_{j\to}\|_q, \tag{2.9}$$

defined by the sum of the ℓ_q norms of the rows of $\boldsymbol{Z}$, for $q \geq 1$. Conditions of recoverability of the solution $\boldsymbol{X}$ of (2.1) by the minimizer of the convex optimization problem (2.8) are established in [15, Theorem 3.1] and [38, Theorem 5.1]. However, there are no conclusive results that demonstrate the advantage of the MMV formulation over the SMV one in the convex relaxation form, as discussed for example in [15, Section D], [38, Section 5.2] and [39, Section 3.2].

These studies make no assumption on the structure of the unknown matrix $\boldsymbol{X}$, except for sparsity of its row support. As shown in section 3, we can write the problem of SAR imaging of direction dependent reflectivities in MMV form, for an unknown matrix $\boldsymbol{X}$ that has almost orthogonal rows. With this assumption on $\boldsymbol{X}$ we prove in the next section that the MMV formulation may have an advantage over SMV.

2.2. Resolution theory. Let us consider the following modification of the linear system (2.1)

$$\boldsymbol{D_W} = \boldsymbol{G}\boldsymbol{X} + \boldsymbol{W}, \tag{2.10}$$

which accounts for data $\boldsymbol{D_W} \in \mathbb{C}^{N_r \times N_v}$ contaminated by the noise matrix $\boldsymbol{W} \in \mathbb{C}^{N_r \times N_v}$. We estimate the unknown $\boldsymbol{X}$ by the minimizer $\boldsymbol{X}^\varepsilon$ of the convex problem

$$\mathscr{P}_{1,2}^\varepsilon: \text{ minimize } \|\boldsymbol{Z}\|_{1,2} \text{ subject to } \|\boldsymbol{G}\boldsymbol{Z} - \boldsymbol{D_W}\|_F \leq \varepsilon, \tag{2.11}$$

where $\|\cdot\|_F$ is the Frobenius norm and ε is a chosen tolerance, satisfying

$$\|\boldsymbol{W}\|_F < \varepsilon. \tag{2.12}$$

Denote by $\mathcal{S} \subset \{1, \ldots, N_{\boldsymbol{y}}\}$ the set of indexes of the nonzero rows of $\boldsymbol{X}$, and suppose that its cardinality $|\mathcal{S}|$ is small with respect to $N_{\boldsymbol{y}}$. We call $\mathcal{S}$ the row support of $\boldsymbol{X}$ and let $\Omega_{\mathcal{S}} = \{\boldsymbol{y}_q, \ q \in \mathcal{S}\}$ be the set of locations in Ω that support the unknowns modeled by $\boldsymbol{X}$. With our notation convention, the rows of $\boldsymbol{X}$ are

denoted by $\boldsymbol{x}_{q\rightarrow}$, and the columns of the sensing matrix $\boldsymbol{G}$ are denoted by $\boldsymbol{g}_q$. We use these vectors to define the "multiple view interaction coefficient"

$$\mathscr{I}_{N_v} = \max_{1 \leq j \leq N_{\boldsymbol{y}}} \sup_{\boldsymbol{v}_\rightarrow \in \mathbb{C}^{1 \times N_v}} \sum_{q \in \mathcal{S} \backslash \{n(j)\}} |\mu(\boldsymbol{g}_j, \boldsymbol{g}_q)| |\mu(\boldsymbol{v}_\rightarrow, \boldsymbol{x}_{q\rightarrow})|, \tag{2.13}$$

which is a measure of how the unknowns influence each other in imaging. The terms in (2.13) involve the correlation of the columns of $\boldsymbol{G}$,

$$\mu(\boldsymbol{g}_j, \boldsymbol{g}_q) = \langle \boldsymbol{g}_j, \boldsymbol{g}_q \rangle, \quad 1 \leq j, q \leq N_{\boldsymbol{y}}, \tag{2.14}$$

where $\langle \boldsymbol{g}_j, \boldsymbol{g}_q \rangle = \boldsymbol{g}_j^\star \boldsymbol{g}_q$ is the Hermitian inner product, and $\star$ denotes complex conjugate and transpose. These columns are normalized, so that

$$\|\boldsymbol{g}_j\|_2 = \langle \boldsymbol{g}_j, \boldsymbol{g}_j \rangle^{1/2} = 1, \quad 1 \leq j \leq N_{\boldsymbol{y}}, \tag{2.15}$$

and we suppose that

$$|\mu(\boldsymbol{g}_j, \boldsymbol{g}_q)| < 1, \quad \forall\, j \neq q,\ 1 \leq j, q \leq N_{\boldsymbol{y}}. \tag{2.16}$$

This assumption holds in the SAR imaging application and it allows us to quantify the distance between the points in Ω using the semimetric

$$\mathfrak{D} : \{1, \ldots, N_{\boldsymbol{y}}\} \times \{1, \ldots, N_{\boldsymbol{y}}\} \rightarrow [0, 1], \quad \mathfrak{D}(j, q) = 1 - |\mu(\boldsymbol{g}_j, \boldsymbol{g}_q)|. \tag{2.17}$$

This semimetric was introduced in [7, Section 4] for the related problem of sensor array imaging of sparse scenes, using ℓ_1 minimization. We will see in section 3 that $|\mu(\boldsymbol{g}_j, \boldsymbol{g}_q)|$ is approximately a function of $\boldsymbol{y}_j - \boldsymbol{y}_q$, which peaks at the origin i.e., for $\boldsymbol{y}_j = \boldsymbol{y}_q$, and decreases monotonically in the vicinity of the peak. Thus, points at small distance with respect to $\mathfrak{D}$ are also close in the Euclidian distance.

We use the semimetric $\mathfrak{D}$ in definition (2.13) to select the closest[‡] point to $\boldsymbol{y}_j$ in $\Omega_\mathcal{S}$, indexed by $n(j)$. In an abuse of notation, we also let $\mu(\cdot, \cdot)$ be the correlation of the rows of $\boldsymbol{X}$ with $\boldsymbol{v}_\rightarrow$, defined by

$$\mu(\boldsymbol{v}_\rightarrow, \boldsymbol{x}_{q\rightarrow}) = \frac{\langle \boldsymbol{v}_\rightarrow, \boldsymbol{x}_{q\rightarrow} \rangle}{\|\boldsymbol{v}_\rightarrow\|_2 \|\boldsymbol{x}_{q\rightarrow}\|_2}, \tag{2.18}$$

where $\langle \boldsymbol{v}_\rightarrow, \boldsymbol{x}_{q\rightarrow} \rangle = \boldsymbol{v}_\rightarrow \boldsymbol{x}_{q\rightarrow}^\star$ is the Hermitian inner product of row vectors and $\|\cdot\|_2$ is the induced ℓ_2 norm.

Note that the correlation (2.18) has absolute value equal to 1 in the SMV setting, where $N_v = 1$ and $\boldsymbol{v}_\rightarrow$ and $\boldsymbol{x}_{k\rightarrow}$ are complex numbers. Then, (2.13) reduces to $\mathscr{I}_1$, the interaction coefficient used in [7, Section 4] to quantify the quality of imaging reconstructions with ℓ_1 optimization. Here we consider $N_v > 1$, and note that since $|\mu(\boldsymbol{v}_\rightarrow, \boldsymbol{x}_{k\rightarrow})| \leq 1$, the multiple view interaction coefficient satisfies $\mathscr{I}_{N_v} \leq \mathscr{I}_1$.

The results in sections 2.2.1–2.2.2 show that the smaller $\mathscr{I}_{N_v}$ and noise level ε are, the better the reconstruction $\boldsymbol{X}^\varepsilon$ of $\boldsymbol{X}$. We do not pursue the question of exact recovery, but seek instead to estimate the neighborhood of the support $\mathcal{S}$ of $\boldsymbol{X}$, where the largest entries in $\boldsymbol{X}^\varepsilon$ lie. The size of this neighborhood may be thought of as the resolution limit. If there are only a few unknowns supported at far apart points, then $\mathscr{I}_{N_v} \leq \mathscr{I}_1 \ll 1$, and the neighborhood will be very small, depending on ε. Moreover, the reconstruction $\boldsymbol{X}^\varepsilon$ will be almost exact. Otherwise, $\mathscr{I}_1$ will not be small but, depending on the rows of $\boldsymbol{X}$, the multiple interaction coefficient may satisfy $\mathscr{I}_{N_v} \ll \mathscr{I}_1$. If this is the case, then the results of our analysis show that there is an advantage in using the MMV methodology.

2.2.1. Estimation of the support of $\boldsymbol{X}$. The next theorem, proved in section 5.2, quantifies the support of the reconstruction $\boldsymbol{X}^\varepsilon$ in terms of $\mathscr{I}_{N_v}$ and the noise level ε.

THEOREM 2.1. *Consider the matrix $\boldsymbol{W}^\varepsilon = \boldsymbol{G}(\boldsymbol{X}^\varepsilon - \boldsymbol{X})$, defined in terms of the unknown solution $\boldsymbol{X}$ and its reconstruction $\boldsymbol{X}^\varepsilon$, the minimizer of (2.11). This matrix cannot be computed but is guaranteed to satisfy the bound*

$$\|\boldsymbol{W}^\varepsilon\|_F \leq 2\varepsilon. \tag{2.19}$$

[‡]If there are more than one such points, we just pick one and let $n(j)$ be its index.

Suppose that there exists $r \in (0,1)$ so that $2\mathscr{I}_{N_v} < r < 1$, and define the set

$$\mathfrak{B}_r(\mathcal{S}) = \{1 \leq j \leq N_{\boldsymbol{y}} \text{ such that } \exists q \in \mathcal{S} \text{ satisfying } \mathfrak{D}(j,q) < r\},$$

called the r–vicinity of $\mathcal{S}$ with respect to the semimetric $\mathfrak{D}$. If we decompose the reconstruction in two parts

$$\boldsymbol{X}^{\varepsilon} = \boldsymbol{X}^{\varepsilon,r} + \boldsymbol{E}^{\varepsilon,r}, \tag{2.20}$$

where $\boldsymbol{X}^{\varepsilon,r}$ has row support in $\mathfrak{B}_r(\mathcal{S})$ and $\boldsymbol{E}^{\varepsilon,r}$ has row support in $\{1,\dots,N_{\boldsymbol{y}}\} \setminus \mathfrak{B}_r(\mathcal{S})$, we have

$$\|\boldsymbol{E}^{\varepsilon,r}\|_{1,2} \leq \frac{2\mathscr{I}_{N_v}}{r}\|\boldsymbol{X}^{\varepsilon}\|_{1,2} + \frac{1}{r}\big\|\big(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon}\big)_{\mathcal{S}\rightarrow}\big\|_{1,2}, \tag{2.21}$$

where $\boldsymbol{G}^{\star} \in \mathbb{C}^{N_{\boldsymbol{y}} \times N_{\boldsymbol{r}}}$ is the Hermitian adjoint of $\boldsymbol{G}$ and $\big(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon}\big)_{\mathcal{S}\rightarrow} \in \mathbb{C}^{|\mathcal{S}| \times N_v}$ is the restriction of the matrix $\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon}$ to the rows indexed by the entries in $\mathcal{S}$.

We may think of the matrix $\boldsymbol{E}^{\varepsilon,r}$ as an error in the reconstruction, because its rows are supported away from $\mathcal{S}$. The theorem says that this error depends on the multiple view interaction coefficient. In the special case $N_v = 1$ (i.e., the SMV formulation) and for $\varepsilon = 0$, the result is precisely [7, Theorem 4.1], which bounds the error in terms of the interaction coefficient $\mathscr{I}_1$. Here we have N_v noisy data sets, and in the case $\mathscr{I}_{N_v} \ll \mathscr{I}_1$, we expect a better estimate of the support $\mathcal{S}$ with MMV than with SMV, because we can choose a smaller r–vicinity of $\mathcal{S}$ that supports the largest entries in the reconstruction $\boldsymbol{X}^{\varepsilon}$.

The noise level ε appears in the last term of (2.21). With the definition of the norm $\|\cdot\|_{1,2}$, the normalization of the columns of $\boldsymbol{G}$ and the bound (2.19), we conclude that this term is at most of order ε,

$$\big\|\big(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon}\big)_{\mathcal{S}\rightarrow}\big\|_{1,2} = \sum_{j\in\mathcal{S}}\|\boldsymbol{g}_j^{\star}\boldsymbol{W}^{\varepsilon}\|_2 = \sum_{j\in\mathcal{S}}\left[\sum_{v=1}^{N_v}|\boldsymbol{g}_j^{\star}\mathbf{w}_v^{\varepsilon}|^2\right]^{1/2} \leq \sum_{j\in\mathcal{S}}\left[\sum_{v=1}^{N_v}\|\mathbf{w}_v^{\varepsilon}\|_2^2\right]^{1/2} = |\mathcal{S}|\|\boldsymbol{W}^{\varepsilon}\|_F \leq 2\varepsilon|\mathcal{S}|,$$

where $\mathbf{w}_v^{\varepsilon}$ are the columns of $\boldsymbol{W}^{\varepsilon}$. This is a pessimistic estimate. In the numerical simulations we found that $\big\|\big(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon}\big)_{\mathcal{S}\rightarrow}\big\|_{1,2}$ is typically much smaller than $2\varepsilon|\mathcal{S}|$.

2.2.2. Quantitative estimation of $\boldsymbol{X}$. Now that we have quantified the error $\boldsymbol{E}^{\varepsilon,r}$ in Theorem 2.1, it remains to study the approximation of the unknown $\boldsymbol{X}$ by $\boldsymbol{X}^{\varepsilon,r} = \boldsymbol{X}^{\varepsilon} - \boldsymbol{E}^{\varepsilon,r}$, which is supported in a set $\mathcal{S}^{\varepsilon} \subset \mathfrak{B}_r(\mathcal{S})$. Because $\mathcal{S}$ and $\mathcal{S}^{\varepsilon}$ are different sets in general, we decompose $\boldsymbol{X}^{\varepsilon,r}$ in two parts: one supported in $\mathcal{S}$ that we compare with $\boldsymbol{X}$, and a residual. We describe first the decomposition and then state the result.

Let us denote by $\boldsymbol{G}_{\mathcal{S}} = (\boldsymbol{g}_j)_{j\in\mathcal{S}}$ the $N_{\boldsymbol{r}} \times |\mathcal{S}|$ matrix obtained by restricting the columns of $\boldsymbol{G}$ to the indexes in $\mathcal{S}$. Suppose that $\boldsymbol{G}_{\mathcal{S}}$ has linearly independent columns, as otherwise it is impossible to recover $\boldsymbol{X}$ even with noiseless data, and introduce its pseudoinverse

$$\boldsymbol{G}_{\mathcal{S}}^{\dagger} = (\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}_{\mathcal{S}})^{-1}\boldsymbol{G}_{\mathcal{S}}^{*}. \tag{2.22}$$

We decompose $\boldsymbol{X}^{\varepsilon,r}$ in two parts

$$\boldsymbol{X}^{\varepsilon,r} = \boldsymbol{\mathfrak{X}}^{\varepsilon,r} + \boldsymbol{\mathcal{E}}^{\varepsilon,r}, \tag{2.23}$$

where $\boldsymbol{\mathfrak{X}}^{\varepsilon,r}$ has row support in $\mathcal{S}$ and its restriction to the rows indexed by $\mathcal{S}$ satisfies

$$\boldsymbol{\mathfrak{X}}_{\mathcal{S}\rightarrow}^{\varepsilon,r} = \boldsymbol{G}_{\mathcal{S}}^{\dagger}\boldsymbol{G}\boldsymbol{X}^{\varepsilon,r}. \tag{2.24}$$

This definition gives that

$$\boldsymbol{G}_{\mathcal{S}}^{\dagger}\boldsymbol{G}\boldsymbol{X}^{\varepsilon,r} = (\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}_{\mathcal{S}})^{-1}\boldsymbol{G}_{\mathcal{S}}^{*}\Big(\boldsymbol{G}_{\mathcal{S}}\boldsymbol{\mathfrak{X}}_{\mathcal{S}\rightarrow}^{\varepsilon,r} + \boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r}\Big) = \boldsymbol{\mathfrak{X}}_{\mathcal{S}\rightarrow}^{\varepsilon,r} + (\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}_{\mathcal{S}})^{-1}\boldsymbol{G}_{\mathcal{S}}^{*}\boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r},$$

so the residual $\boldsymbol{\mathcal{E}}^{\varepsilon,r}$ satisfies

$$\boldsymbol{G}_{\mathcal{S}}^{*}\boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r} = 0. \tag{2.25}$$

5

That is to say, the columns of $G\mathcal{E}^{\varepsilon,r}$ are orthogonal to the range of $G_{\mathcal{S}}$. Note that $\mathcal{E}^{\varepsilon,r}$ has row support in $\mathcal{S} \cup \mathcal{S}^\varepsilon$. If $\mathcal{S}^\varepsilon$ were the same as $\mathcal{S}$, then (2.25) would imply that $\mathcal{E}^{\varepsilon,r} = 0$. Thus, we may think of $\mathcal{E}^{\varepsilon,r}$ as a residual that accounts for $X^{\varepsilon,r}$ not having the exact support $\mathcal{S}$.

The matrix $\mathfrak{X}^{\varepsilon,r}$, which is row supported in $\mathcal{S}$ by definition, is compared to the unknown X in the next theorem. However, since $\mathfrak{X}^{\varepsilon,r}$ cannot be computed directly, we also need to relate it to the reconstruction $X^{\varepsilon,r}$ supported near $\mathcal{S}$, in the set $\mathcal{S}^\varepsilon$. The next theorem relates $\mathfrak{X}^{\varepsilon,r}$ to the "effective matrix" $\overline{X^{\varepsilon,r}}$ obtained by local aggregation of the rows of $X^{\varepsilon,r}$. To define this aggregation, let us decompose the set $\mathcal{S}^\varepsilon$ in $|\mathcal{S}|$ disjoint parts[§], each corresponding to a point in $\mathcal{S}$,

$$\mathcal{S}^\varepsilon = \bigcup_{j \in \mathcal{S}} \mathfrak{S}_j, \quad \mathfrak{S}_j = \{q \in \mathcal{S}^\varepsilon \text{ such that } \mathfrak{D}(q,j) \le \mathfrak{D}(q,j'), \ \forall \ j' \in \mathcal{S}\}, \quad j \in \mathcal{S}. \tag{2.26}$$

The effective matrix $\overline{X^{\varepsilon,r}} \in \mathbb{C}^{N_{\boldsymbol{y}} \times N_v}$ is defined entry wise by

$$\overline{X_{j,v}^{\varepsilon,r}} = \begin{cases} \displaystyle\sum_{l \in \mathfrak{S}_j} X_{l,v}^{\varepsilon,r} \mu(\boldsymbol{g}_j, \boldsymbol{g}_l), & \text{if } j \in \mathcal{S}, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } 1 \le j \le N_{\boldsymbol{y}}, \ 1 \le v \le N_v. \tag{2.27}$$

Note that $\mu(\boldsymbol{g}_j, \boldsymbol{g}_l)$ is complex valued. However, when the set $\mathfrak{S}_j$ is small, so that $\mathfrak{D}(j,l) \ll 1$ for $l \in \mathfrak{S}_j$, $\mu(\boldsymbol{g}_j, \boldsymbol{g}_l) \approx 1$ and (2.27) is approximately the local sum of the entries of $X^{\varepsilon,r}$.

THEOREM 2.2. *Let $X^{\varepsilon,r}$ and $\mathfrak{X}^{\varepsilon,r}$ be be defined as in (2.20) and (2.23). Then, $\mathfrak{X}^{\varepsilon,r}$ approximates the unknown X with the error estimate*

$$\|\mathfrak{X}^{\varepsilon,r} - X\|_{1,2} \le \frac{2\mathscr{I}_{N_v}}{r}\|X^\varepsilon\|_{1,2} + \frac{3}{r}\|(G^\star W^\varepsilon)_{\mathcal{S}\to}\|_{1,2}, \tag{2.28}$$

where W^ε is the matrix defined in Theorem 2.1. Moreover, if the support of $X^{\varepsilon,r}$ is decomposed in $|\mathcal{S}|$ disjoint parts as in (2.26), $\mathfrak{X}^{\varepsilon,r}$ is approximated by the effective matrix $\overline{X^{\varepsilon,r}}$ with entries (2.27), as follows

$$(1 - \mathscr{I}_1)\|\mathfrak{X}^{\varepsilon,r} - \overline{X^{\varepsilon,r}}\|_{1,1} \le 2\mathscr{I}_1 \|X^{\varepsilon,r}\|_{1,1}. \tag{2.29}$$

This theorem, proved in section 5.3, shows that under the same conditions as in Theorem 2.1, which guarantee that the minimizer X^ε of (2.11) has support near $\mathcal{S}$, the matrix $\mathfrak{X}^{\varepsilon,r}$ is a good approximation of the unknown X. If the right hand side in (2.29) is small, which occurs when the points in $\Omega_{\mathcal{S}}$ are well separated [7], then $\mathfrak{X}^{\varepsilon,r}$ is close to the effective matrix $\overline{X^{\varepsilon,r}}$. The bound in (2.29) involves the interaction coefficient $\mathscr{I}_1$ for the SMV problem, which in our setting may be much larger than $\mathscr{I}_{N_v}$. This means that while a small $\mathscr{I}_{N_v}$ may guarantee a good estimate of the support of X per Theorem 2.1, the estimate of the entries in X by the local sum in (2.27), which is close to the sum of the entries of X^ε, may not be accurate when the unknowns are supported at nearby points.

2.2.3. Matrices X with orthogonal rows. We now show that if the unknown matrix X has orthogonal rows, then the multiple view interaction coefficient $\mathscr{I}_{N_v}$ may be much smaller than the interaction coefficient $\mathscr{I}_1$. In light of the results in Theorem 2.1, this means that the MMV approach can give improved estimates of the support $\mathcal{S}$ of X, under less stringent conditions on the separation between the points in $\mathcal{S}$, than in the SMV formulation.

PROPOSITION 2.3. *Suppose that the unknown matrix $X \in \mathbb{C}^{N_{\boldsymbol{y}} \times N_v}$ has row support in the set $\mathcal{S}$ with cardinality $1 < |\mathcal{S}| \le N_v$, and that its nonzero rows are orthogonal. Then, the multiple view interaction coefficient (2.13) is given by*

$$\mathscr{I}_{N_v} = \max_{1 \le j \le N_{\boldsymbol{y}}} \sqrt{\sum_{q \in \mathcal{S}\setminus\{n(j)\}} |\mu(\boldsymbol{g}_j, \boldsymbol{g}_q)|^2}. \tag{2.30}$$

[§]If more than one point in $\mathcal{S}$ is at the same distance to $\boldsymbol{y}_q$, with $q \in \mathcal{S}^\varepsilon$, we pick any one of them so that $\mathfrak{S}_j$ are disjoints.

This proposition, proved in section 5.4, gives a simpler expression of $\mathscr{I}_{N_v}$, which we can compare with the interaction coefficient

$$\mathscr{I}_1 = \max_{1 \le j \le N_{\boldsymbol{y}}} \sum_{q \in \mathcal{S} \setminus \{n(j)\}} |\mu(\boldsymbol{g}_j, \boldsymbol{g}_q)|, \tag{2.31}$$

to understand when $\mathscr{I}_{N_v} \ll \mathscr{I}_1$. For this purpose, let us define the vector $\boldsymbol{\gamma}^{(j)} \in \mathbb{R}^{1 \times (|\mathcal{S}|-1)}$ with entries $|\mu(\boldsymbol{g}_j, \boldsymbol{g}_q)|$, for $q \in \mathcal{S} \setminus \{n(j)\}$, which is a set with cardinality $|\mathcal{S}| - 1$, and rewrite (2.30) and (2.31) as

$$\mathscr{I}_{N_v} = \max_{1 \le j \le N_{\boldsymbol{y}}} \|\boldsymbol{\gamma}^{(j)}\|_2, \qquad \mathscr{I}_1 = \max_{1 \le j \le N_{\boldsymbol{y}}} \|\boldsymbol{\gamma}^{(j)}\|_1, \tag{2.32}$$

using the ℓ_2 and ℓ_1 vector norms. Suppose that the maximizer in the definition of $\mathscr{I}_{N_v}$ is at index $j = m$. Basic vector norm inequalities give the general relation

$$\mathscr{I}_{N_v} = \|\boldsymbol{\gamma}^{(m)}\|_2 \le \|\boldsymbol{\gamma}^{(m)}\|_1 \le \mathscr{I}_1,$$

which is nothing new than was discussed previously. However, if we assume further that the entries in $\boldsymbol{\gamma}^{(m)}$ are of the same order, meaning that there exist positive numbers β^- and β^+, ordered as $\beta^- \le \beta^+$ and satisfying $\beta^+/\beta^- = O(1)$, such that

$$\beta^- \le |\mu(\boldsymbol{g}_m, \boldsymbol{g}_q)| \le \beta^+, \quad \forall\, q \in \mathcal{S} \setminus \{n(m)\}, \tag{2.33}$$

then we have

$$\mathscr{I}_{N_v} \le \beta^+ \sqrt{|\mathcal{S}| - 1} = \frac{\beta^+ \left[\beta^- (|\mathcal{S}| - 1)\right]}{\beta^- \sqrt{|\mathcal{S}| - 1}} \le \frac{\beta^+ \|\boldsymbol{\gamma}^{(m)}\|_1}{\beta^- \sqrt{|\mathcal{S}| - 1}} \le \frac{\beta^+ \mathscr{I}_1}{\beta^- \sqrt{|\mathcal{S}| - 1}} = O\left(\frac{\mathscr{I}_1}{\sqrt{|\mathcal{S}| - 1}}\right). \tag{2.34}$$

Recalling the discussion below definition (2.17) of the semimetric $\mathfrak{D}$ and that $|\mu(\boldsymbol{g}_m, \boldsymbol{g}_q)| = 1 - \mathfrak{D}(m, q)$, we can interpret the condition (2.33) as having points in the support $\Omega_{\mathcal{S}}$ of the unknown evenly distributed, at similar spacing in the imaging region Ω. If this condition holds, then the multiple view interaction coefficient is smaller than the single view coefficient $\mathscr{I}_1$, by order $\sqrt{|\mathcal{S}|}$. We illustrate this further with numerical simulations in section 3.

2.2.4. Clusters of unknowns. For general distributions of the points in $\Omega_{\mathcal{S}}$, the multiple view interaction coefficient $\mathscr{I}_{N_v}$ may be large, so we cannot conclude from Theorems 2.1 and 2.2 that the reconstruction $\boldsymbol{X}^\varepsilon$ approximates $\boldsymbol{X}$. Here we consider an extension of the results to the case of points in $\Omega_{\mathcal{S}}$ that are clustered around a few locations, indexed by the elements in the set $\mathcal{C} \subset \{1, \ldots, N_{\boldsymbol{y}}\}$ of cardinality $|\mathcal{C}| \ll |\mathcal{S}|$.

We decompose the set $\mathcal{S}$ in $|\mathcal{C}|$ disjoint parts, called "cluster sets",

$$\mathcal{S} = \bigcup_{j \in \mathcal{C}} \mathscr{S}_j, \quad \mathscr{S}_j = \{q \in \mathcal{S} \text{ such that } \mathfrak{D}(q, j) < \mathfrak{D}(q, j'), \ \forall j' \in \mathcal{C}, j' \ne j\}, \quad j \in \mathcal{C}, \tag{2.35}$$

where $j \in \mathcal{C}$ indexes the centers of the clusters and we assume that the distance (measured by $\mathfrak{D}$) between these centers is larger than the radius of each cluster set. We also define the "effective cluster matrix" $\overline{\boldsymbol{X}}$ by aggregating the rows of $\boldsymbol{X}$ over the cluster sets. Explicitly, $\overline{\boldsymbol{X}} \in \mathbb{C}^{N_{\boldsymbol{y}} \times N_v}$ is row supported in $\mathcal{C}$, with entries

$$\overline{\boldsymbol{X}}_{j,v} = \begin{cases} \displaystyle\sum_{l \in \mathscr{S}_j} \boldsymbol{X}_{l,v}\, \mu(\boldsymbol{g}_j, \boldsymbol{g}_l), & j \in \mathcal{C}, \\ 0, & \text{otherwise}, \end{cases} \quad \text{for } 1 \le j \le N_{\boldsymbol{y}}, \ 1 \le v \le N_v. \tag{2.36}$$

If the radius of the cluster sets $\mathscr{S}_j$ is small, we have $\mu(\boldsymbol{g}_l, \boldsymbol{g}_j) \approx 1$ and the aggregation in (2.36) is approximately the sum of the rows of $\boldsymbol{X}$. Our goal in this section is to study in what sense the reconstruction $\boldsymbol{X}^\varepsilon$, the solution of the optimization problem (2.11), approximates $\overline{\boldsymbol{X}}$.

7

Let $\boldsymbol{G}_{\mathcal{C}} = (\boldsymbol{g}_j)_{j \in \mathcal{C}}$ be the restriction of the sensing matrix to the columns indexed in $\mathcal{C}$, and suppose that $\boldsymbol{G}_{\mathcal{C}}$ has independent columns, so that we can define its pseudoinverse by

$$\boldsymbol{G}_{\mathcal{C}}^{\dagger} = (\boldsymbol{G}_{\mathcal{C}}^{\star} \boldsymbol{G}_{\mathcal{C}})^{-1} \boldsymbol{G}_{\mathcal{C}}^{\star}. \tag{2.37}$$

Then, we can rewrite the data model (2.10) in terms of a new unknown matrix $\boldsymbol{U} \in \mathbb{C}^{N_{\boldsymbol{y}} \times N_v}$ for the cluster, with row support in $\mathcal{C}$. This matrix is not the same as the effective cluster matrix $\overline{\boldsymbol{X}}$, but is defined by the projection of $\boldsymbol{X}$ on $\mathcal{C}$, such that its restriction to the rows indexed by $\mathcal{C}$ satisfies

$$\boldsymbol{U}_{\mathcal{C} \to} = \boldsymbol{G}_{\mathcal{C}}^{\dagger} \boldsymbol{G} \boldsymbol{X}. \tag{2.38}$$

We write (2.10) in the form

$$\boldsymbol{G} \boldsymbol{U} + \boldsymbol{\mathcal{W}} = \boldsymbol{D}_{\boldsymbol{W}}, \tag{2.39}$$

where $\boldsymbol{R} = \boldsymbol{X} - \boldsymbol{U}$ is the residual after the projection, and

$$\boldsymbol{\mathcal{W}} = \boldsymbol{W} + \boldsymbol{G}\boldsymbol{R}, \tag{2.40}$$

is the new "noise matrix", consisting of the actual noise $\boldsymbol{W}$ and the systematic error term $\boldsymbol{G}\boldsymbol{R}$. The next lemma, proved in section 5.5, quantifies this error in terms of the radius of the row support of $\boldsymbol{X}$ around the cluster centers indexed by $\mathcal{C}$.

LEMMA 2.4. *Suppose that each cluster set $\mathscr{S}_j$ in the decomposition (2.35) is supported within a ball of radius r_c around the point $j \in \mathcal{C}$, with respect to the semimetric $\mathfrak{D}$, for all $j \in \mathcal{C}$. Then,*

$$\|\boldsymbol{G}\boldsymbol{R}\|_F \leq \sqrt{2r_c}\|\boldsymbol{X}^T\|_{2,1}, \tag{2.41}$$

where the index T denotes the transpose.

Assuming that r_c is small, we introduce a new "noise level" ε, so that

$$\|\boldsymbol{\mathcal{W}}\|_F = \|\boldsymbol{W} + \boldsymbol{G}\boldsymbol{R}\|_F < \varepsilon. \tag{2.42}$$

We also define the multiple view interaction coefficient as in (2.13), with the set $\mathcal{S}$ replaced by the smaller set $\mathcal{C}$ and the row vectors $\boldsymbol{x}_{q \to}$ replaced by the rows $\boldsymbol{u}_{q \to}$ of $\boldsymbol{U}$,

$$\mathscr{I}_{N_v}^{\boldsymbol{U}} = \max_{1 \leq j \leq N_{\boldsymbol{y}}} \sup_{\boldsymbol{v}_{\to} \in \mathbb{C}^{1 \times N_v}} \sum_{q \in \mathcal{C} \setminus \{n(j)\}} |\mu(\boldsymbol{g}_j, \boldsymbol{g}_q)| |\mu(\boldsymbol{v}_{\to}, \boldsymbol{u}_{q \to})|. \tag{2.43}$$

The next theorem, proved in section 5.5, is the extension of Theorem 2.1 for the cluster.

THEOREM 2.5. *Let $\boldsymbol{X}^{\varepsilon}$ be the minimizer of (2.11), with ε chosen large enough to satisfy (2.42). Decompose it in two parts*

$$\boldsymbol{X}^{\varepsilon} = \boldsymbol{U}^{\varepsilon,r} + \boldsymbol{E}^{\varepsilon,r}, \tag{2.44}$$

where $\boldsymbol{U}^{\varepsilon,r}$ is row supported in the set $\mathfrak{B}_r(\mathcal{C})$, the r vicinity of $\mathcal{C}$ with respect to the semimetric $\mathfrak{D}$, and $\boldsymbol{E}^{\varepsilon,r}$ is the error supported in the complement $\{1, \ldots, N_{\boldsymbol{y}}\} \setminus \mathfrak{B}_r(\mathcal{C})$. This satisfies the estimate

$$\|\boldsymbol{E}^{\varepsilon,r}\|_{1,2} \leq \frac{2\mathscr{I}_{N_v}^{\boldsymbol{U}}}{r} \|\boldsymbol{X}^{\varepsilon}\|_{1,2} + \frac{1}{r} \left\| \left(\boldsymbol{G}^{\star} \boldsymbol{W}^{\varepsilon}\right)_{\mathcal{C} \to} \right\|_{1,2}, \tag{2.45}$$

with $\boldsymbol{W}^{\varepsilon} = \boldsymbol{G}(\boldsymbol{X}^{\varepsilon} - \boldsymbol{X})$ defined as in Theorem 2.1.

This result quantifies the error $\boldsymbol{E}^{\varepsilon,r}$ that is supported away from the center points of the clusters in terms of the multiple view interaction coefficient (2.43). If these center points are sufficiently far apart, then $\mathscr{I}_{N_v}^{\boldsymbol{U}} \ll \mathscr{I}_{N_v}$ and the estimate (2.45) is an improvement over that in Theorem 2.1. Note that the same matrix $\boldsymbol{W}^{\varepsilon}$ enters these estimates, except that in (2.45) the row restriction is on the smaller set $\mathcal{C}$. However, the "noise level" ε may be larger than in Theorem 2.1, in order to accommodate the systematic error estimated

in (2.41). If the radius r_c is so small that this estimate is of the order of $\|\boldsymbol{W}\|_F$, then the theorem says that the optimization approximates the set $\mathcal{C}$ with the same accuracy as if we had single unknowns located at $\boldsymbol{y}_j$, for $j \in \mathcal{C}$, and not clusters of points.

Finally, we state the analogue of Theorem 2.2 for the cluster, proved in section 5.5. This quantifies the approximation of the effective cluster matrix (2.36) by the part $\boldsymbol{U}^{\varepsilon,r}$ of the minimizer $\boldsymbol{X}^{\varepsilon}$.

THEOREM 2.6. *Let $\boldsymbol{U}^{\varepsilon,r}$ be decomposed in two parts*

$$\boldsymbol{U}^{\varepsilon,r} = \mathscr{U}^{\varepsilon,r} + \boldsymbol{\mathcal{E}}^{\varepsilon,r}, \tag{2.46}$$

where $\mathscr{U}^{\varepsilon,r} \in \mathbb{C}^{N_{\boldsymbol{y}} \times N_v}$ is the matrix with row support in $\mathcal{C}$, satisfying

$$\mathscr{U}_{\mathcal{C}\to}^{\varepsilon,r} = \boldsymbol{G}_{\mathcal{C}}^{\dagger}\boldsymbol{G}\boldsymbol{U}^{\varepsilon,r}. \tag{2.47}$$

This matrix approximates the cluster unknown matrix $\boldsymbol{U}$ as

$$\|\mathscr{U}^{\varepsilon,r} - \boldsymbol{U}\|_{1,2} \le \frac{2\mathscr{I}_{N_v}^{\boldsymbol{U}}}{r}\|\boldsymbol{X}^{\varepsilon}\|_{1,2} + \frac{3}{r}\|\left(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon}\right)_{\mathcal{C}\to}\|_{1,2}. \tag{2.48}$$

Moreover, if we let $\overline{\boldsymbol{U}^{\varepsilon,r}}$ be the effective reconstruction matrix with rows supported in $\mathcal{C}$, the analogue of (2.27) with $\boldsymbol{X}^{\varepsilon,r}$ replaced by $\boldsymbol{U}^{\varepsilon,r}$ and $\mathcal{S}$ replaced by $\mathcal{C}$, then we have

$$(1 - \mathscr{I}_1^{\boldsymbol{U}})\|\mathscr{U}^{\varepsilon,r} - \overline{\boldsymbol{U}^{\varepsilon,r}}\|_{1,1} \le 2\mathscr{I}_1^{\boldsymbol{U}}\|\boldsymbol{U}^{\varepsilon,r}\|_{1,1}. \tag{2.49}$$

Finally, the matrix $\boldsymbol{U}$ is related to the effective cluster matrix $\overline{\boldsymbol{X}}$ defined in (2.36) by

$$(1 - \mathscr{I}_1^{\boldsymbol{U}})\|\boldsymbol{U} - \overline{\boldsymbol{X}}\|_{1,1} \le 2\mathscr{I}_1^{\boldsymbol{U}}\|\boldsymbol{X}\|_{1,1}. \tag{2.50}$$

This result says that if the center points of the clusters are sufficiently far apart, so the single view interaction coefficient $\mathscr{I}_1^{\boldsymbol{U}}$ is smaller than $\mathscr{I}_1$, the estimates (2.49)–(2.50) are an improvement over those in Theorem 2.2. However, the reconstruction is not an approximation of $\boldsymbol{X}$, but of the effective matrix $\overline{\boldsymbol{X}}$.

3. SAR imaging of direction dependent reflectivity. In this section we consider the application of SAR imaging of direction dependent reflectivities. We begin with the data model in section 3.1, and then derive in section 3.2 the linear system (2.1). The discussion in these two sections is very similar to that in [8], so we keep it short and give only the information that is needed to connect to the theory in section 2.2. We explore in section 3.3 the condition of orthogonality of the rows of $\boldsymbol{X}$, assumed in Proposition 2.3, and use numerical simulations in section 3.4 to illustrate the theoretical results.

3.1. The SAR data model. Consider the set-up illustrated in Figure 3.1, where we display a piece of the synthetic aperture spanned by the moving transmit-receive antenna, called a sub-aperture. We approximate the sub-aperture by a line segment along the unit vector $\boldsymbol{\tau}$, with center at location $\overline{\boldsymbol{r}}$, and length a. The imaging region Ω lies on a plane surface, and is centered at location $\overline{\boldsymbol{y}}$, at distance $\overline{L} = |\overline{\boldsymbol{r}} - \overline{\boldsymbol{y}}|$ from the center of the aperture. The antenna probes the imaging region by emitting periodically the signal $f(t)$, and measuring the back-scattered waves. The waves propagate much faster than the antenna, so we assume that the emission and reception occur at the same location. The antenna moves by a small increment $\Delta\boldsymbol{r} = \frac{a}{(N_{\boldsymbol{r}}-1)}\boldsymbol{\tau}$ between two emissions, so the measurements are made at the $N_{\boldsymbol{r}}$ locations

$$\boldsymbol{r}_j = \overline{\boldsymbol{r}} - \frac{a\boldsymbol{\tau}}{2} + (j-1) * \Delta\boldsymbol{r}, \quad j = 1, \dots, N_{\boldsymbol{r}}. \tag{3.1}$$

In the single scattering (Born) approximation, and neglecting for now polarization effects, the scattered wave field measured at $\boldsymbol{r}_j$ is given by

$$p(\boldsymbol{r}_j, t; \overline{\boldsymbol{r}}, \overline{\omega}) = \int \frac{d\omega}{2\pi} e^{-i\omega t} \widehat{f}(\omega) k^2(\omega) \sum_{q=1}^{N_{\boldsymbol{y}}} \rho_q(\overline{\boldsymbol{r}}, \overline{\omega}) \frac{\exp\left[2ik(\omega)|\boldsymbol{r}_j - \boldsymbol{y}_q|\right]}{\left(4\pi|\boldsymbol{r}_j - \boldsymbol{y}_q|\right)^2}. \tag{3.2}$$
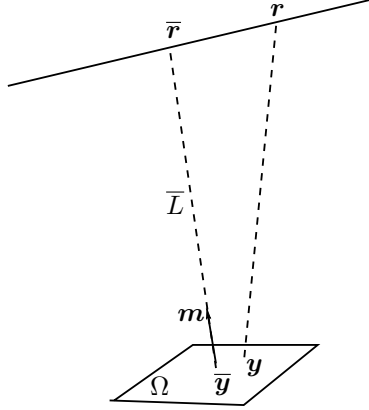
FIG. 3.1. *Setup for SAR imaging using a linear synthetic aperture centered at $\overline{r}$, at distance $\overline{L}$ from the center $\overline{y}$ of the imaging region Ω. The antenna locations r span the aperture of length a, and y denotes a point in Ω. The unit vector $m = (\overline{r} - \overline{y})/\overline{L}$ pointing from $\overline{y}$ to $\overline{r}$ defines the range direction.*

Here the hat denotes Fourier transform with respect to time, ω is the frequency, $\rho_q(\overline{r}, \omega)$ is the reflectivity[¶] of the scatterer at $y_q \in \Omega$, as viewed from the aperture centered at $\overline{r}$ and at the central frequency $\overline{\omega}$ of $f(t)$. The propagation of the waves between the antenna location r_j and y_q is modeled with the Green's function for Helmoltz's equation in the medium with constant wave speed c, and the wavenumber is $k(\omega) = \omega/c$.

In SAR imaging, the wave-field (3.2) is convolved with the time reversed emitted pulse, delayed by the round trip travel time of the waves between the antenna and the center point $\overline{y}$ in Ω. This data processing is called down-ramping [20] and we denote the result by

$$d(r_j, t; \overline{r}, \overline{\omega}) = p(r_j, t; \overline{r}, \overline{\omega}) \star_t f^\star(-t - 2|r_j - \overline{y}|/c), \tag{3.3}$$

where $f^\star$ denotes the complex conjugate of f. Writing the time convolution in terms of the Fourier transform, and letting

$$|\widehat{f}(\omega)|^2 = \widehat{\varphi}\left(\frac{\omega - \overline{\omega}}{b}\right), \tag{3.4}$$

with $\widehat{\varphi}$ a non-negative, smooth function of dimensionless argument and support in the interval $(-\pi, \pi)$, so that b denotes the bandwidth of the signal, we obtain the data model

$$d(r_j, t; \overline{r}, \overline{\omega}) = \int \frac{dw}{2\pi} e^{-i(\overline{\omega}+w)t} \widehat{\varphi}\left(\frac{w}{b}\right) k^2(\overline{\omega}+w) \sum_{q=1}^{N_y} \rho_q(\overline{r}, \overline{\omega}) \frac{\exp\left[2ik(\overline{\omega}+w)\left(|r_j - y_q| - |r_j - \overline{y}|\right)\right]}{\left(4\pi|r_j - y_q|\right)^2}. \tag{3.5}$$

3.1.1. Scaling regime and simplification of data model. We define the unit vector $m = (\overline{r} - \overline{y})/\overline{L}$ which determines the so-called range direction in imaging, and introduce the length scales

$$Y = \sup_{y \in \Omega} |(y - \overline{y}) \cdot m|, \qquad Y^\perp = \sup_{y \in \Omega} |\mathbb{P}(y - \overline{y})|, \tag{3.6}$$

where $\mathbb{P} = \mathbf{I} - mm^T$ is the orthogonal projection in the cross-range plane, orthogonal to m. These length scales quantify the range and cross-range size of the imaging region Ω.

We consider as in [8] a Fresnel diffraction regime, with $\overline{L} \gg a > Y^\perp \gg \overline{\lambda}$ and Fresnel numbers satisfying

$$\frac{a^2}{\overline{\lambda}\overline{L}} \gtrsim \frac{aY^\perp}{\overline{\lambda}\overline{L}} \gtrsim \frac{(Y^\perp)^2}{\overline{\lambda}\overline{L}} \gtrsim 1, \tag{3.7}$$

[¶]The reflectivity is assumed slowly changing so it can be approximated by a constant over this sub-aperture and bandwidth.

where $\overline{\lambda} = 2\pi/\overline{k}$ is the central wavelength and $\overline{k} = k(\overline{\omega}) = \overline{\omega}/c$. The cross-range resolution of SAR imaging is of the order $\overline{\lambda}\overline{L}/a$, so the middle inequality in (3.7) ensures that the cross-range size $Y^{\perp}$ of the imaging region is large with respect to this resolution limit. The other inequalities in (3.7) mean physically that the wave front observed at the sub-aperture or in Ω is not planar. If this were not the case, it would be impossible to localize the scatterers in cross-range.

The range resolution of SAR imaging is determined by the accuracy of travel time estimation from the down-ramped data (3.5). It is of the order c/b, so we let the range scale Y of the imaging region be larger

$$Y \gtrsim c/b. \tag{3.8}$$

Typically $b \ll \overline{\omega}$, especially when imaging frequency dependent reflectivities. To simplify the data model, we assume that b is small enough so that

$$\frac{b}{\overline{\omega}} \frac{a Y^{\perp}}{\overline{\lambda}\,\overline{L}} \ll 1. \tag{3.9}$$

This can be arranged by division of a larger bandwidth in smaller sub-bands. Similarly, we can ensure a synthetic aperture segmentation in sub-aperture sizes a satisfying

$$\frac{a^2 Y}{\overline{\lambda}\,\overline{L}^2} \ll 1, \quad \frac{a^2 Y^{\perp}}{\overline{\lambda}\,\overline{L}^2} \ll 1. \tag{3.10}$$

Under the scaling assumptions (3.7)–(3.10) the data model (3.5) takes the simpler form [8, Section 3.1]

$$d(\boldsymbol{r}_j, t; \overline{\boldsymbol{r}}, \overline{\omega}) \approx \left(\frac{\overline{k}}{4\pi\overline{L}}\right)^2 \sum_{q=1}^{N_{\boldsymbol{y}}} \rho_q(\overline{\boldsymbol{r}}, \overline{\omega})\varphi\left[b\left(t + \frac{2\boldsymbol{m}\cdot\Delta\boldsymbol{y}_q}{c}\right)\right]e^{-i\overline{\omega}t}\times$$
$$\exp\left[-2i\overline{k}\left(\boldsymbol{m}\cdot\Delta\boldsymbol{y}_q + \frac{\Delta\boldsymbol{r}_j\cdot\mathbb{P}\Delta\boldsymbol{y}_q}{\overline{L}} - \frac{\Delta\boldsymbol{y}_q\cdot\mathbb{P}\Delta\boldsymbol{y}_q}{2\overline{L}}\right)\right], \tag{3.11}$$

where we introduced the notation

$$\Delta\boldsymbol{r}_j = \boldsymbol{r}_j - \overline{\boldsymbol{r}}, \quad \Delta\boldsymbol{y}_q = \boldsymbol{y}_q - \overline{\boldsymbol{y}}.$$

Note that by varying t in (3.11), we can limit the sum to the set of points with range coordinates satisfying $\boldsymbol{m}\cdot\Delta\boldsymbol{y}_q = -t + O(c/b)$. This set is called a range bin in the SAR literature [20].

We consider a single range bin, for fixed time $t = \overline{t}$, and study the estimation in the cross-range direction of the reflectivity, for the single frequency sub-band centered at $\overline{\omega}$. Then, the model (3.11) takes the form

$$\mathscr{D}_j(\overline{\boldsymbol{r}}) = \sum_{q=1}^{N_{\boldsymbol{y}}} \frac{\exp\left[-2i\overline{k}\frac{\Delta\boldsymbol{r}_j\cdot\mathbb{P}\Delta\boldsymbol{y}_q}{\overline{L}}\right]}{\sqrt{N_r}} \mathscr{X}_q(\overline{\boldsymbol{r}}), \tag{3.12}$$

with the notation

$$\mathscr{X}_q(\overline{\boldsymbol{r}}) = \rho_q(\overline{\boldsymbol{r}}, \omega)\sqrt{N_r}\varphi\left[b\left(\overline{t} + \frac{2\boldsymbol{m}\cdot\Delta\boldsymbol{y}_q}{c}\right)\right]\exp\left[-2i\overline{k}\left(\boldsymbol{m}\cdot\Delta\boldsymbol{y}_q - \frac{\Delta\boldsymbol{y}_q\cdot\mathbb{P}\Delta\boldsymbol{y}_q}{2\overline{L}}\right)\right], \tag{3.13}$$

and

$$\mathscr{D}_j(\overline{\boldsymbol{r}}) = d(\boldsymbol{r}_j, t; \overline{\boldsymbol{r}}, \overline{\omega})e^{i\overline{k}\,\overline{t}}\left(\frac{4\pi\overline{L}}{\overline{k}}\right)^2. \tag{3.14}$$

Here we suppressed the constant variables in the notation and assume that all the $N_{\boldsymbol{y}}$ points in the sum are in the same range bin, determined by $\overline{t}$.

3.2. The MMV formulation. The multiple views in the MMV formulation correspond to different sub-apertures, with centers at $\overline{\boldsymbol{r}}_v$, for $v = 1, \dots, N_v$. The noiseless data model for the v–th view is given by (3.12), with $\overline{\boldsymbol{r}}$ replaced by $\overline{\boldsymbol{r}}_v$ and $\overline{L}$ replaced by $\overline{L}_v = |\overline{\boldsymbol{r}}_v - \overline{\boldsymbol{y}}|$. We also let $\boldsymbol{m}_v = |\overline{\boldsymbol{r}}_v - \overline{\boldsymbol{y}}|/\overline{L}_v$ and denote the unit tangent vector to the aperture by $\boldsymbol{\tau}_v$.

To fuse all the views in a linear system of the form (2.1), we suppose as in [8] that

$$\max_{1 \le v \le N_v, 1 \le q \le N_{\boldsymbol{y}}} \frac{a}{\overline{\lambda}} \left| \left(\frac{\boldsymbol{\tau}_v \cdot \mathbb{P}_v}{\overline{L}_v} - \frac{\boldsymbol{\tau}_1 \cdot \mathbb{P}_1}{\overline{L}_1} \right) \Delta \boldsymbol{y}_q \right| \ll 1, \qquad \max_{1 \le v \le N_v, 1 \le q \le N_{\boldsymbol{y}}} \frac{b}{c} \left| (\boldsymbol{m}_v - \boldsymbol{m}_1) \cdot \Delta \boldsymbol{y}_q \right| \ll 1. \quad (3.15)$$

Physically, this means that the imaging points remain within the classic SAR resolution limits for all the views. With (3.15) we derive from (3.12) the linear system (2.1), for matrices $\boldsymbol{D}$, $\boldsymbol{X}$ and $\boldsymbol{G}$ with entries

$$D_{j,v} = \mathscr{D}_j(\overline{\boldsymbol{r}}_v), \quad X_{q,v} = \mathscr{X}_q(\overline{\boldsymbol{r}}_j), \quad G_{j,q} = \frac{1}{\sqrt{N_r}} \exp\left[-2i\overline{k} \frac{\Delta \boldsymbol{r}_j \cdot \mathbb{P}_1 \Delta \boldsymbol{y}_q}{\overline{L}_1} \right], \quad (3.16)$$

where $\mathbb{P}_1 = I - \boldsymbol{m}_1 \boldsymbol{m}_1^T$.

Note that the sensing matrix $\boldsymbol{G}$ is defined relative to the first sub-aperture, the line segment centered at $\overline{\boldsymbol{r}}_1$, along the unit vector $\boldsymbol{\tau}_1$. Its columns $\boldsymbol{g}_q$ have norm one, as assumed in (2.15), and their correlation

$$\mu(\boldsymbol{g}_q, \boldsymbol{g}_l) = \sum_{j=1}^{N_r} G_{j,q}^\star G_{j,l} = \frac{1}{N_r} \sum_{j=1}^{N_r} \exp\left[-2i\overline{k} \frac{\Delta \boldsymbol{r}_j \cdot \mathbb{P}_1 (\boldsymbol{y}_q - \boldsymbol{y}_l)}{\overline{L}_1} \right] \quad (3.17)$$

is a function of $\boldsymbol{y}_q - \boldsymbol{y}_l$, as stated below equation (2.17). We can approximate further this correlation by replacing the sum with the integral over the sub-aperture,

$$\mu(\boldsymbol{g}_q, \boldsymbol{g}_l) \approx \frac{1}{a} \int_{-a/2}^{a/2} dr \, \exp\left[-2i\overline{k}r \frac{\boldsymbol{\tau}_1 \cdot \mathbb{P}_1 (\boldsymbol{y}_q - \boldsymbol{y}_l)}{\overline{L}_1} \right] = \text{sinc}\left(\frac{\overline{k}a\boldsymbol{\tau}_1 \cdot \mathbb{P}_1 (\boldsymbol{y}_q - \boldsymbol{y}_l)}{\overline{L}_1} \right). \quad (3.18)$$

This attains its maximum, equal to 1, when $q = l$, and satisfies $|\mu(\boldsymbol{g}_q, \boldsymbol{g}_l) < 1$ for all $q \neq l$, as assumed in (2.16). Moreover, $|\mu(\boldsymbol{g}_q, \boldsymbol{g}_l)|$ decays monotonically in the vicinity of its peak, so we can relate the Euclidian distance between the points to the semimetric $\mathfrak{D}(q, l)$, as pointed out below equation (2.17).

3.3. Orthogonality of the rows. We now study under which conditions the rows $\boldsymbol{x}_{q\to}$ of $\boldsymbol{X}$ are orthogonal, as assumed in section 2.2.3.

Let us introduce the notation

$$\xi_{q,v} = \rho_q(\overline{\boldsymbol{r}}_v, \omega) \sqrt{N_r} \, \varphi\left[b\left(\overline{t} + \frac{2\boldsymbol{m}_v \cdot \Delta \boldsymbol{y}_q}{c} \right) \right] \approx \rho_q(\overline{\boldsymbol{r}}_v, \omega) \sqrt{N_r} \, \varphi\left[b\left(\overline{t} + \frac{2\boldsymbol{m}_1 \cdot \Delta \boldsymbol{y}_q}{c} \right) \right], \quad (3.19)$$

with the approximation due to assumption (3.15) and the smoothness of φ. Definitions (3.13), (3.16) give

$$\left| \mu(\boldsymbol{x}_{q\to}, \boldsymbol{x}_{l\to}) \right| = \frac{|\langle \boldsymbol{x}_{q\to}, \boldsymbol{x}_{l\to} \rangle|}{\|\boldsymbol{x}_{q\to}\|_2 \|\boldsymbol{x}_{l\to}\|_2} \approx \frac{\left| \sum_{v=1}^{N_v} \xi_{q,v} \xi_{l,v}^\star \exp\left[2i\overline{k} \boldsymbol{m}_v \cdot (\boldsymbol{y}_q - \boldsymbol{y}_l) \right] \right|}{\sqrt{\sum_{v=1}^{N_v} |\xi_{q,v}|^2} \sqrt{\sum_{v=1}^{N_v} |\xi_{l,v}|^2}}, \quad (3.20)$$

where the quadratic terms in the phase in (3.13), which are approximately independent of the view index v by assumption (3.15), do not contribute to the absolute value.

There are various possibilities for achieving a small right hand side in (3.20), depending on how rapidly (3.19) changes with the view index v. Consistent with our approximation of the reflectivity by a constant over each sub-aperture, we assume that $\rho_q(\overline{\boldsymbol{r}}_v, \omega)$ changes slowly with $\overline{\boldsymbol{r}}_v$, on a length scale larger than a. We also suppose that all the sub-apertures have the same direction $\boldsymbol{\tau}_1$ and they overlap, with two consecutive centers separated by a small distance

$$|\boldsymbol{r}_{v+1} - \boldsymbol{r}_v| = \frac{A}{N_v - 1} \ll a, \quad v = 1, \dots, N_v - 1.$$

12

This allows us to approximate the sum over v in (3.20) by the integral over the large aperture $A \gg a$, centered at $\boldsymbol{r}_o$,

$$\left|\mu(\boldsymbol{x}_{q\to}, \boldsymbol{x}_{l\to})\right| \approx \frac{\left|\int_{-A/2}^{A/2} dr\, \psi_{q,l}(r) \exp\left[2i\overline{k}\frac{(\boldsymbol{r}_o + r\boldsymbol{\tau}_1 - \overline{\boldsymbol{y}})}{|\boldsymbol{r}_o + r\boldsymbol{\tau}_1 - \overline{\boldsymbol{y}}|} \cdot (\boldsymbol{y}_q - \boldsymbol{y}_l)\right]\right|}{\|\psi_{q,q}\|_{L_1(-A/2,A/2)}^{1/2}\|\psi_{l,l}\|_{L_1(-A/2,A/2)}^{1/2}}. \tag{3.21}$$

Here we parametrize the large aperture by the arclength $r \in [-A/2, A/2]$, and $\psi_{q,l}(r)$ is the smooth kernel satisfying the interpolation conditions

$$\psi_{q,l}\left(r = \left(\frac{v-1}{N_v - 1} - \frac{1}{2}\right)A\right) = \xi_{q,v}\xi_{l,v}^{\star}, \tag{3.22}$$

at the arclength r corresponding to the v–th sub-aperture center $\overline{\boldsymbol{r}}_v = \boldsymbol{r}_o + \left(\frac{v-1}{N_v-1} - \frac{1}{2}\right)A\boldsymbol{\tau}_1$, for $v = 1, \ldots, N_v$.

PROPOSITION 3.1. *There exists a constant $C_{q,l}$ that depends on how fast the reflectivities at points $\boldsymbol{y}_q$ and $\boldsymbol{y}_l$ change with direction, such that*

$$\left|\mu(\boldsymbol{x}_{q\to}, \boldsymbol{x}_{l\to})\right| \le \min\{1, C_{q,l}/|Q|\}, \quad \text{for } q \ne l, \quad q, l = q, \ldots, N_{\boldsymbol{y}}, \tag{3.23}$$

where

$$Q = 4\pi\frac{A\boldsymbol{\tau}_1 \cdot \mathbb{P}_o(\boldsymbol{y}_q - \boldsymbol{y}_l)}{\overline{\lambda}|\boldsymbol{r}_o - \overline{\boldsymbol{y}}|}, \quad \mathbb{P}_o = \boldsymbol{I} - \boldsymbol{m}_o\boldsymbol{m}_o^T, \quad \boldsymbol{m}_o = \frac{(\boldsymbol{r}_o - \overline{\boldsymbol{y}})}{|\boldsymbol{r}_o - \overline{\boldsymbol{y}}|}. \tag{3.24}$$

This proposition, proved in Appendix A, shows that the correlation of the rows of the unknown matrix $\boldsymbol{X}$ is large for points that are separated in cross-range by at most order $\overline{\lambda}|\boldsymbol{r}_o - \overline{\boldsymbol{y}}|/A$. This length scale is the cross-range resolution of SAR imaging over the large aperture A. It is also the distance at which isotropic scatterers must be separated in order to guarantee unique recovery of their reflectivity with ℓ_1 (SMV) optimization over the large aperture, as follows from [28, 13, 14, 7].

In the linear system (2.1) with matrices (3.16), we use multiple views from sub-apertures of size $a \ll A$. A single view corresponds to an SMV problem, and the condition of unique recovery would be that the points should be much further apart, at distance of order $\overline{\lambda}|r_o - \overline{\boldsymbol{y}}|/a$. In MMV we use the entire large aperture, segmented in N_v smaller sub-apertures.

When the scatterers are approximately isotropic, the constant in (3.23) is $C_{q,l} \approx 2$. In this case there is no need to segment the aperture, so it is natural to ask if the MMV reconstruction is similar to the SMV one, over the large aperture. This is a difficult question, but we can say from the results in section 2.2.3 that MMV will work better$^{\|}$ then SMV over one sub-aperture, because the rows of the unknown matrix $\boldsymbol{X}$ are approximately orthogonal when the points in its support are at distances of order $\overline{\lambda}|r_o - \overline{\boldsymbol{y}}|/A \ll \overline{\lambda}|r_o - \overline{\boldsymbol{y}}|/a$.

When the scatterers have a stronger dependence on direction, the SMV approach over the large aperture does not work well. Aperture segmentation is needed to avoid systematic modeling errors in the optimization. While we may apply the SMV approach for a single sub-aperture, Proposition 3.1 and the results in section 2.2.3 show that the MMV method performs better.

3.4. Numerical results. We present here numerical results that illustrate the theory presented in section 2.2. We begin in section 3.4.1 with a computational assessment of the reduction of the multiple view interaction coefficient with respect to the single view one, in the case of orthogonal rows of the unknown matrix $\boldsymbol{X}$. Then we present in section 3.4.2 imaging results, using the parameters of the X-band GOTCHA SAR data set [1]: The receive-transmit platform moves on a linear aperture $A = 1.5$km at altitude 8km, and with center $\overline{r}_o$ at 7km west of $\overline{\boldsymbol{y}}$. The platform emits and receives signals every meter. The central frequency is 10GHz and since we only present imaging in cross-range, the bandwidth plays no role. The waves propagate at speed $c = 3 \cdot 10^8$m/s.

The data are generated numerically using the single scattering approximation. The additive noise matrix $\boldsymbol{W}$ has mean zero and independent complex Gaussian entries with standard deviation σ given as a percent of the largest entry in $\boldsymbol{D}$. The optimization problem (2.11) is solved using the software package CVX [21].

$^{\|}$ As shown in section 2.2.3, the improvement is dependent on the distribution of the scatterers in the imaging region.
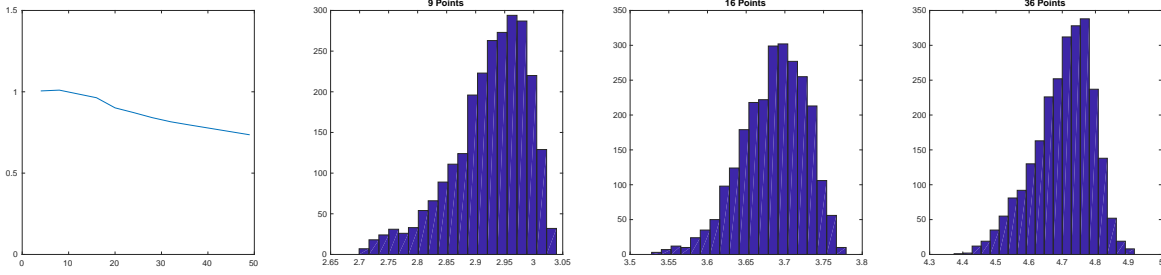
FIG. 3.2. *Left plot: The ratio* $\dfrac{\mathscr{I}_1}{\mathscr{I}_{N_v}\sqrt{|\mathcal{S}|}}$ *vs.* $|S|$ *in the abscissa. The other plots: Histograms of the ratio* $\mathscr{I}_1/\mathscr{I}_{N_v}$ *for* 2500 *realizations of the imaging scene. From left to right* $|\mathcal{S}|$ *equals* 9, 16 *and* 36. *The ordinate shows the number of realizations and the abscissa is the value of* $\mathscr{I}_1/\mathscr{I}_{N_v}$.

3.4.1. Numerical illustration of effects of orthogonality of rows of X. The discussion in section 2.2.3 says that if the points in $\Omega_{\mathcal{S}}$ are distributed evenly in the imaging window Ω, and the rows of X are orthogonal, then the multiple view interaction coefficient $\mathscr{I}_{N_v}$ is smaller than $\mathscr{I}_1$, by a factor of order $\sqrt{|\mathcal{S}|}$. Here we focus attention on imaging in the cross-range direction, so the imaging region is reduced to a line segment. We cannot have a large number of points with similar mutual separation on a line. Nevertheless, we show here that the numerically computed ratio $\mathscr{I}_1/\mathscr{I}_{N_v}$ increases with $|S|$, at a rate that is slightly slower than $\sqrt{|\mathcal{S}|}$.

We display in Figure 3.2 the ratio $\mathscr{I}_1/\mathscr{I}_{N_v}$ computed for imaging scenes with $|S|$ ranging from 4 to 50, and cross-range separation of nearby neighbors chosen randomly, uniformly distributed in the interval $\left[\overline{\lambda}\,\overline{L}_o/A, 3\overline{\lambda}\,\overline{L}_o/A\right]$, where $\overline{L}_o = |\overline{r}_o - \overline{y}|$. The large aperture A is divided in sub-apertures of size $a = A/20$. The rows of X have length 50 and are orthogonal**, to stay within the setting of section 2.2.3.

The left plot in Figure 3.2 shows the ratio $\mathscr{I}_1/\left(\mathscr{I}_{N_v}\sqrt{|\mathcal{S}|}\right)$ computed for one realization of the imaging scene. We note that the increase of $\mathscr{I}_1/\mathscr{I}_{N_v}$ with $|\mathcal{S}|$ is slightly slower than $\sqrt{|\mathcal{S}|}$. The histograms in Figure 3.2, computed for 2500 realizations of the imaging scene, also show that the ratio is slightly less than $\sqrt{|\mathcal{S}|}$.

3.4.2. Imaging results. We begin with a comparison of imaging results obtained with the MMV optimization formulation (2.11) for $N_v = 24$, the SMV formulation for $N_v = 1$, and the conventional SAR image. The latter is given by the superposition of the down-ramped data (3.5), synchronized using the round-trip travel time of the waves from the radar platform to the imaging point

$$I^{SAR}(\boldsymbol{y};\overline{\boldsymbol{r}}) = \sum_{j=1}^{N_r} d\big(\boldsymbol{r}_j, t = 2|\boldsymbol{r}_j - \boldsymbol{y}|/c; \overline{\boldsymbol{r}}, \overline{\omega}\big). \tag{3.25}$$

The superposition may be over the entire aperture centered at $\overline{\boldsymbol{r}} = \overline{\boldsymbol{r}}_o$, in which case $N_r = 1500$, or over a sub-aperture, centered at $\overline{\boldsymbol{r}} = \overline{\boldsymbol{r}}_v$ for $v = 1, \ldots, N_v$, in which case $N_r = 300$. The sub-aperture length is $a = A/6 = 300$m, and the spacing between the sub-apertures is 50m, center to center. The results in Figures 3.3–3.5 are for noiseless data and in Figure (3.6) we consider noise with standard deviation $\sigma = 10\%$.

We display in the left plot of Figure 3.3 the reflectivity to be reconstructed. It consists of 6 small scatterers at cross-range locations spaced by distances of order $\overline{\lambda}\,\overline{L}_o/A$. The conventional SAR image computed over the entire aperture $A = 1.5$km is shown in the middle plot. It does not resolve well the location of the five scatterers that are visible only on about a sixth of A, but it obtains a large peak for the one scatterer with reflectivity that varies less with direction. In the right plot of Figure 3.3 we display the average of the images (3.25) obtained over the N_v sub-apertures. This does have peaks near the scatterers, for example the rightmost scatterer is better seen. However, since the resolution is $\overline{\lambda}\,\overline{L}_o/a$, the image appears blurry at the cross-range scale $\overline{\lambda}\,\overline{L}_o/A$ in the figure.

**As shown in section 3.3, the rows of X are almost orthogonal when its support points are spaced at distances larger than $\overline{\lambda}\,\overline{L}_o/A$. Thus, the ratio $\mathscr{I}_1/\mathscr{I}_{N_v}$ is expected to be roughly like in the figures for any reflectivity function that varies smoothly with the direction of illumination.
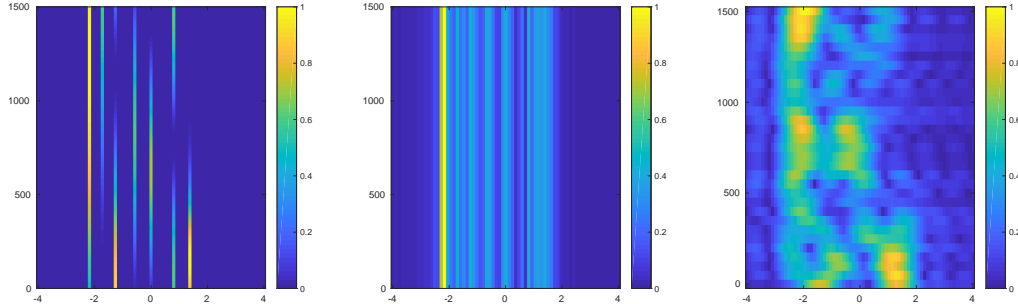
14

Fig. 3.3. *Left plot: Exact reflectivity function as viewed from the location on the flight path (ordinate, in meters) vs. the cross-range location in the imaging scene (abscissa, in units $\overline{\lambda}\overline{L}_o/A$). Middle plot: The conventional SAR image (3.25) calculated over the entire aperture. This assumes that the reflectivity is isotropic (does not change along the ordinate). Right plot: The average of the conventional SAR images (3.25) calculated over the $N_v = 24$ sub-apertures.*
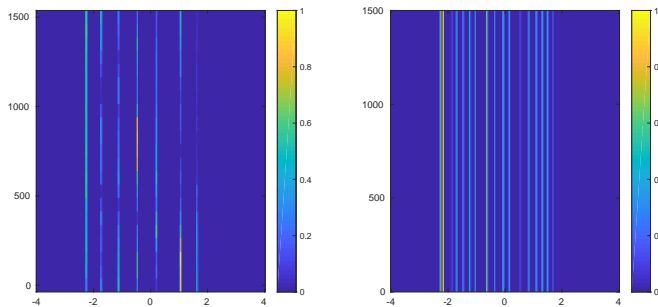

Fig. 3.4. *Left plot: The MMV reconstruction of the reflectivity shown in the left plot in Figure 3.3. Right plot: The SMV reconstruction. The axes are as in Figure 3.3.*

In Figure 3.4 we compare the results obtained with the MMV and SMV approach. We note that the MMV method recovers exactly the support of the scatterers, whereas the SMV method has many spurious peaks. This result is an illustration of the conclusion reached at the end of section 2.2.2, which says that MMV may give a better estimate of the support of the scatterers. However, the estimate of the value of the reflectivity is not accurate, unless the scatterers are further apart.

In Figure 3.5 we consider reflectivities that vary more rapidly over directions, and compare the effect of the size of the sub-aperture on the quality of the reconstructions with the MMV approach. The images show that the best reconstruction is for $a = 70$m, which corresponds roughly with the scale of variation of the true reflectivity in the top plot. For the smaller aperture $a = 40$m (left, bottom plot) the reconstructed support is close but not exact, whereas for the larger aperture $a = 100$m (right, bottom plot) the image has spurious peaks caused by the systematic error due to the reflectivity varying on a smaller scale than the sub-aperture. Thus, we conclude that in order to image successfully direction dependent reflectivities, it is necessary to have a good estimate of their scale of variation, so that the aperture is properly segmented.

In Figure 3.5 we display the effect of additive noise with standard deviation $\sigma = 10\%$ on the MMV reconstruction of the reflectivity, for sub-aperture size $a = 70$m. We note that for such noise the support of the reconstruction is basically unchanged and the values of the reflectivity are only slightly different. Naturally, at higher noise levels, the reconstruction will be worse.

4. SAR imaging with polarization diverse measurements. In this section we describe briefly the application of the MMV methodology to SAR imaging with polarization diverse measurements. We begin in section 4.1 with the derivation of the data model (2.1) used in the MMV method, and then show numerical results in section 4.2.
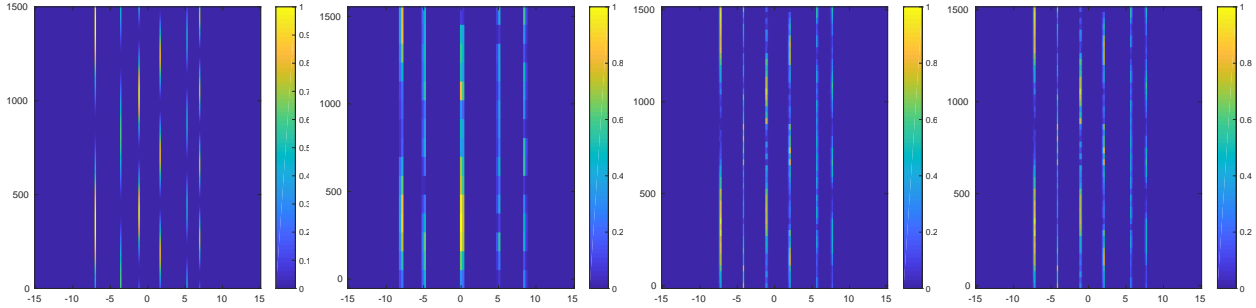
FIG. 3.5. *Left plot: Exact reflectivity function as viewed from the location on the flight path (ordinate, in meters), vs. the cross-range location in the imaging scene (abscissa, in units $\overline{\lambda}\,\overline{L}_o/A$). Other plots: The MMV reconstruction for apertures $a = 50m$, $70m$ and $100m$, from left to right.*
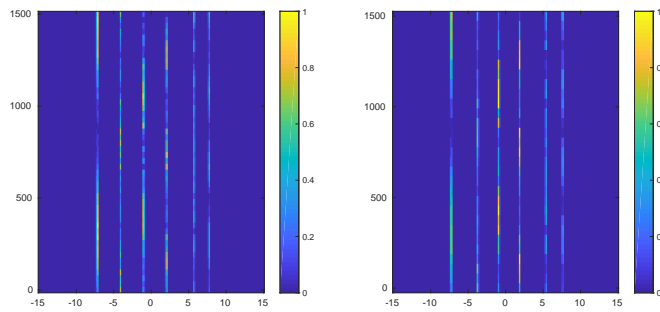

FIG. 3.6. *Comparison of MMV reconstructions with noiseless data (left plot) and noisy data (right plot). The noise is additive, complex Gaussian, with mean zero independent entries and standard deviation $\sigma = 10\%$ of the largest entry in $\boldsymbol{D}$. The axes are as in Figure 3.3.*

4.1. Data model. We consider as in the previous section a collection of $|\mathcal{S}|$ scatterers at locations $\boldsymbol{y}_q \in \Omega$, for $q \in \mathcal{S}$. The scatterers are penetrable, with volume smaller than $\overline{\lambda}^3$ by a factor $\alpha \ll 1$, so that the scattered electric field at the SAR platform can be modeled by [3]

$$\mathscr{E}(\boldsymbol{r}_j, t; \overline{\boldsymbol{r}}, \boldsymbol{f}) = \int \frac{d\omega}{2\pi} e^{-i\omega t} ik^3(\omega) \sqrt{\frac{\mu}{\epsilon}} \sum_{q \in \mathcal{S}} \widehat{\mathscr{G}}(\omega, \boldsymbol{r}_j, \boldsymbol{y}_q) \boldsymbol{\rho}_q(\overline{\boldsymbol{r}}) \widehat{\mathscr{G}}(\omega, \boldsymbol{y}_q, \boldsymbol{r}_j) \widehat{\boldsymbol{f}}(\omega) + O(\alpha^4), \tag{4.1}$$

where $\overline{\lambda}$ is the central wavelength and μ and ϵ are the magnetic permeability and the electric permittivity in the medium. These define the wave speed $c = 1/\sqrt{\mu\epsilon}$ and the wavenumber $k(\omega) = \omega/c$. The scatterers are represented in (4.1) by their center location $\boldsymbol{y}_q$ and their reflectivity tensor assumed constant over the sub-aperture centered at $\overline{\boldsymbol{r}}$,

$$\boldsymbol{\rho}_q(\overline{\boldsymbol{r}}) = \alpha^3 \left(\frac{\epsilon_q}{\epsilon} - 1 \right) \boldsymbol{M}_q(\overline{\boldsymbol{r}}), \tag{4.2}$$

where ϵ_q is the electric permittivity in the scatterer and $\boldsymbol{M}_q$ is its α–independent polarization tensor. We refer to [2] for details on $\boldsymbol{M}_q$, which depends on the shape of the scatterer. Here we assume that it is a real valued 3×3 symmetric matrix. Since we consider a fixed central frequency $\overline{\omega}$, we suppress in the notation the dependence of $\boldsymbol{\rho}_q$ on $\overline{\omega}$. We also neglect the variation of the magnetic permeability in the scatterer, although this can be taken into account, as shown in [2].

The wave propagation from the radar platform to the scatterers and back is modeled in (4.1) by the dyadic Green's tensor

$$\widehat{\mathscr{G}}(\omega, \boldsymbol{r}, \boldsymbol{y}) = \left(\boldsymbol{I} + \frac{\nabla\nabla^T}{k^2(\omega)} \right) \frac{\exp[ik(\omega)|\boldsymbol{r} - \boldsymbol{y}|]}{4\pi|\boldsymbol{r} - \boldsymbol{y}|}, \tag{4.3}$$
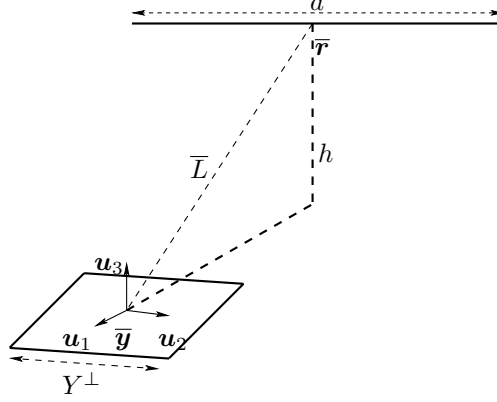
16

FIG. 4.1. *Geometry of the data acquisition. The radar platform flies at elevation h from the plane surface containing the imaging region Ω, centered at $\overline{y}$. The distance $\overline{L}$ from the center $\overline{r}$ of the aperture to $\overline{y}$ is order h. The drawing is not up to scale, as the aperture a and side $Y^\perp$ of the imaging region are much smaller than $\overline{L}$.*

where $\boldsymbol{I}$ is the 3×3 identity matrix. The wave excitation is modeled by the vector $\widehat{\boldsymbol{f}}(\omega)$. To avoid a lengthy discussion[††] we assume that the radar platform can emit and receive all possible polarizations, so that we have access to the 3×3 frequency dependent data matrix

$$\widehat{\mathscr{D}}(\boldsymbol{r}_j, \omega; \overline{\boldsymbol{r}}) \approx \sum_{q=1}^{N_{\boldsymbol{y}}} \widehat{\mathscr{G}}(\omega, \boldsymbol{r}_j, \boldsymbol{y}_q) \boldsymbol{\rho}_q(\overline{\boldsymbol{r}}) \widehat{\mathscr{G}}(\omega, \boldsymbol{y}_q, \boldsymbol{r}_j), \tag{4.4}$$

with the approximation due to the neglected $O(\alpha^4)$ residual. Note that in (4.4) we sum over all the $N_{\boldsymbol{y}}$ points in the imaging region, with the convention that $\boldsymbol{\rho}_q = 0$ for $q \notin \mathcal{S}$.

As in the previous section, we focus attention on imaging in the cross-range direction, so it is sufficient to consider a single frequency, equal to the central one $\overline{\omega}$. The wave number at this frequency is denoted by $\overline{k}$, as in the previous section.

The sub-aperture centered at $\overline{\boldsymbol{r}}$ is linear, of length a, like before, and we assume for simplicity that it is at constant altitude h, as shown in Figure 4.1. We let $\boldsymbol{u}_3$ be the unit vector in the vertical direction, and introduce the unit vector $\boldsymbol{u}_1 = \boldsymbol{\tau} \times \boldsymbol{u}_3$, where $\boldsymbol{\tau}$ is the unit tangent to the aperture, orthogonal to $\boldsymbol{u}_3$. The imaging region Ω is in the plane spanned by $\boldsymbol{u}_1$ and $\boldsymbol{\tau}$. We are interested in its cross-section in the direction of the aperture, which is the cross-range interval centered at $\overline{\boldsymbol{y}}$, of length $Y^\perp$.

In the system of coordinates with center at $\overline{\boldsymbol{y}}$ and orthonormal basis $\{\boldsymbol{u}_j\}_{1 \le j \le 3}$, with $\boldsymbol{u}_2 = \boldsymbol{\tau}$, we have

$$\boldsymbol{r} = r_1 \boldsymbol{u}_1 + r_2 \boldsymbol{u}_2 + h \boldsymbol{u}_3, \quad \boldsymbol{y} = y_2 \boldsymbol{u}_2, \tag{4.5}$$

for all $\boldsymbol{r}$ in the aperture and $\boldsymbol{y}$ in the cross-range imaging interval. We also represent the symmetric 3×3 reflectivity tensor $\boldsymbol{\rho}_q(\overline{\boldsymbol{r}})$ by the 1×6 row-vector formed with the entries in its upper-tridiagonal part

$$\boldsymbol{\rho}_{q\rightarrow} = (\rho_{q,11}, \rho_{q,22}, \rho_{q,33}, \rho_{q,12}, \rho_{q,13}, \rho_{q,23}), \quad \rho_{q,jl} = \boldsymbol{u}_j^T \boldsymbol{\rho}_q \boldsymbol{u}_l. \tag{4.6}$$

The scaling regime is as in the previous section, with length scales ordered as $\overline{\lambda} \ll Y^\perp \lesssim a \ll h$ and

$$\overline{L} = |\overline{\boldsymbol{r}}| = O(h), \quad |r_j| = O(\overline{L}), \quad j = 1, 2. \tag{4.7}$$

The Green tensor (4.3) has the following approximation in this regime

$$\begin{aligned}
\widehat{\mathscr{G}}(\omega, \boldsymbol{r}_j, \boldsymbol{y}_q) &\approx \frac{\exp[i\overline{k}|\boldsymbol{r}_j - \boldsymbol{y}_q|]}{4\pi\overline{L}} \left[\boldsymbol{I} - \frac{(\boldsymbol{r}_j - \boldsymbol{y}_q)(\boldsymbol{r}_j - \boldsymbol{y}_q)^T}{|\boldsymbol{r}_j - \boldsymbol{y}_q|^2} \right] \\
&\approx \frac{\exp[i\overline{k}|\boldsymbol{r}_j - \boldsymbol{y}_q|]}{4\pi\overline{L}} \begin{pmatrix} 1 - \eta_1^2 & -\eta_1\eta_2 & -\eta_1\beta \\ -\eta_1\eta_2 & 1 - \eta_2^2 & -\eta_2\beta \\ -\eta_1\beta & -\eta_2\beta & 1 - \beta^2 \end{pmatrix}, \quad \eta_j = \overline{r}_j/\overline{L}, \;\; j = 1, 2, \;\; \beta = h/\overline{L}.
\end{aligned} \tag{4.8}$$

[††]In fact, only the transverse components of the electric field, in the plane orthogonal to the range direction $\overline{\boldsymbol{r}} - \overline{\boldsymbol{y}}$, play a role in the end, as discussed at the end of this section.
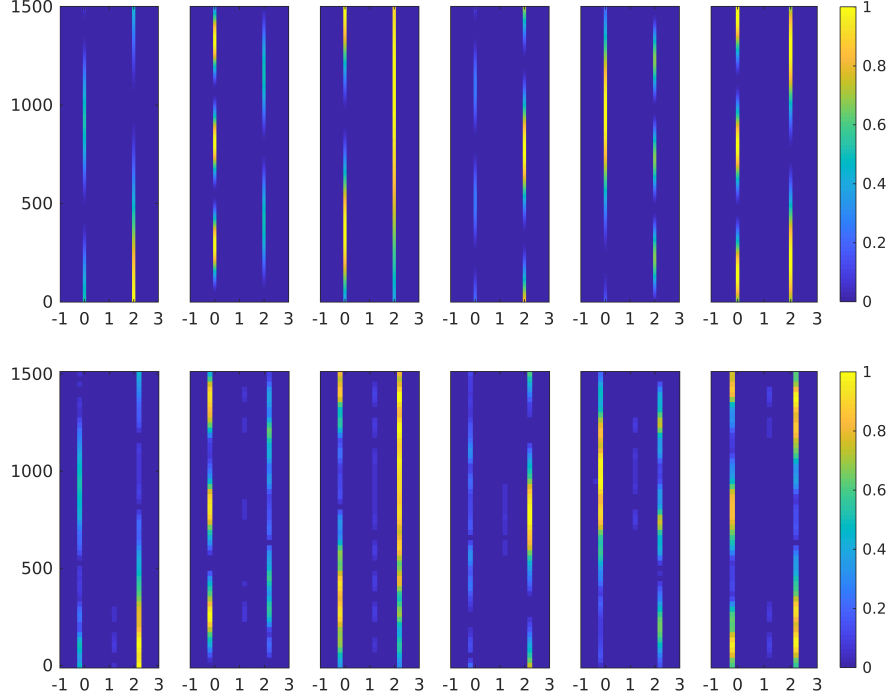
FIG. 4.2. *Top line: From left to right we display all six components of the row vector $\boldsymbol{x}_{q\to}(\overline{\boldsymbol{r}})$ defined in (4.10), as a function of location along the aperture (the ordinate in meters) and cross-range location indexed by q in the imaging region (the abscissa, in units of $\overline{\lambda}\,\overline{L}_o/A$). Bottom line: The MMV reconstruction.*

Substituting it in (4.4), and representing the symmetric matrix $(4\pi\overline{L})^2/\sqrt{N_r}\,\widehat{\mathscr{D}}(\boldsymbol{r}_j,\overline{\omega};\overline{\boldsymbol{r}})$ by the 1×6 row vector formed with the entries in its upper triangular part, we obtain the data model

$$\boldsymbol{d}_{j\to}(\overline{\boldsymbol{r}}) = \sum_{q=1}^{N_{\boldsymbol{y}}} \frac{\exp[2i\overline{k}|\boldsymbol{r}_j - \boldsymbol{y}_q|]}{\sqrt{N_r}} \boldsymbol{\rho}_{q\to}(\overline{\boldsymbol{r}})\boldsymbol{\Gamma}(\overline{\boldsymbol{r}}), \quad j = 1, \ldots, N_r, \tag{4.9}$$

with $\overline{\boldsymbol{r}} = \overline{r}_1\boldsymbol{u}_1 + \overline{r}_2\boldsymbol{u}_2 + h\boldsymbol{u}_3$ and constant matrix $\boldsymbol{\Gamma}(\overline{\boldsymbol{r}})$ given in Appendix B. This is a linear system of form (2.1), for $N_v = 6$, data matrix $\boldsymbol{D} \in \mathbb{C}^{N_r\times 6}$ with rows $\boldsymbol{d}_{j\to}$, unknown matrix $\boldsymbol{X} \in \mathbb{C}^{N_{\boldsymbol{y}}\times 6}$ with rows

$$\boldsymbol{x}_{q\to} = \boldsymbol{\rho}_{q\to}\boldsymbol{\Gamma}, \tag{4.10}$$

and sensing matrix $\boldsymbol{G}$ with normalized columns $\boldsymbol{g}_q = \frac{1}{\sqrt{N_r}}\left(\exp[2i\overline{k}|\boldsymbol{r}_1 - \boldsymbol{y}_q|], \ldots, \exp[2i\overline{k}|\boldsymbol{r}_{N_r} - \boldsymbol{y}_q|]\right)^T$.

The system (4.9) is for a single sub-aperture. More sub-apertures, centered at $\overline{\boldsymbol{r}}_v$, can be taken into account as explained in the previous section, with the only difference being that instead of having a scalar unknown, we now have the unknown 1×6 row vector $\boldsymbol{\rho}_q(\overline{\boldsymbol{r}}_v)\boldsymbol{\Gamma}(\overline{\boldsymbol{r}}_v)$. The linear system that fuses the data from all the sub-apertures is obtained as in section 3.2, and the unknown matrix $\boldsymbol{X}$ has six times more columns than in the acoustic case.

Note that the approximation (4.8) of the Green's tensor $\widehat{\mathscr{G}}(\omega, \boldsymbol{r}_j, \boldsymbol{y}_q)$ for the sub-aperture centered at $\overline{\boldsymbol{r}}$ has the one dimensional null space span$\{\overline{\boldsymbol{r}}\}$. This implies that the matrix $\boldsymbol{\Gamma}(\overline{\boldsymbol{r}})$ is also singular, so we cannot determine uniquely the reflectivity vectors $\boldsymbol{\rho}_{q\to}$ from equation (4.10). To be more explicit, we can represent the reflectivity tensor $\boldsymbol{\rho}_q$ in (4.4) in the sub-aperture dependent orthonormal basis $\{\boldsymbol{v}_j\}_{j=1,2,3}$ of eigenvectors of the matrix in (4.8), with $\boldsymbol{v}_3 = \overline{\boldsymbol{r}}/|\overline{\boldsymbol{r}}|$. Then, we obtain that the components $\{\boldsymbol{v}_j^T\boldsymbol{\rho}_q\boldsymbol{v}_3\}_{j=1,2,3}$ play no role in the data model (4.4), so we can only estimate $(\boldsymbol{v}_j^T\boldsymbol{\rho}_q\boldsymbol{v}_l)_{j,l=1,2}$. This ambiguity is due to the scaling relation $a/|\overline{\boldsymbol{r}}| \ll 1$ and it implies that only the transverse components of the electric field are needed in imaging, as the longitudinal component along $\boldsymbol{v}_3$ adds no information. If the reflectivity tensor does not
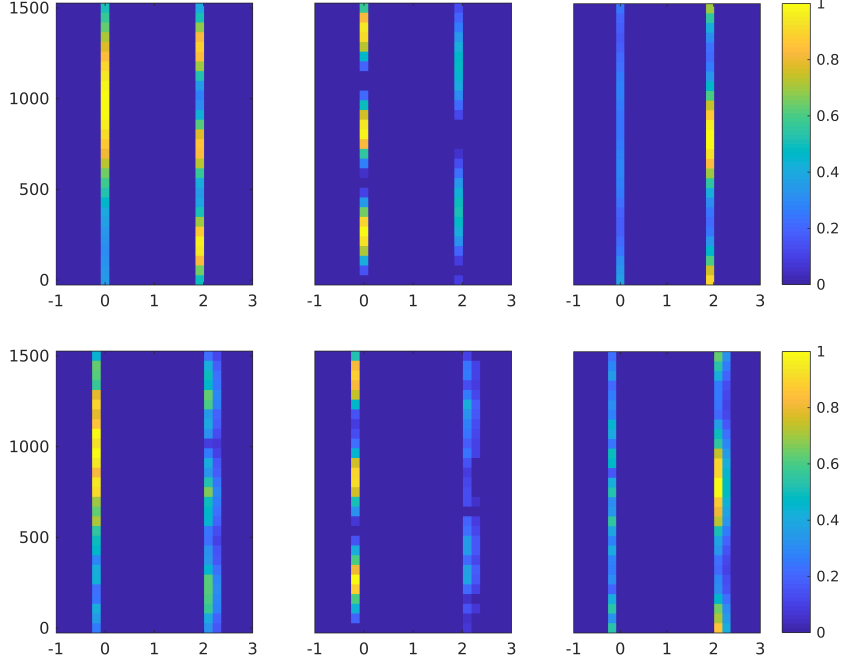
18

FIG. 4.3. *Top line: The components $\boldsymbol{v}_j^T \boldsymbol{\rho}_q \boldsymbol{v}_l$ of the reflectivity matrix, for $j = l = 1$ (left plot), $j = l = 2$ (middle plot) and $j = 1, l = 2$ (right plot). The orthonormal basis $(\boldsymbol{v}_j)_{j=1,2,3}$ depends on the center location $\overline{\boldsymbol{r}}$ of the sub-aperture (the ordinate in meters). The abscissa is the cross-range location indexed by q, in units of $\overline{\lambda}\,\overline{L}_o/A$. Bottom line: The reconstruction.*

change over directions, or it changes slowly, then the ambiguity can be overcome by taking into consideration the multiple sub-apertures, because $\overline{\boldsymbol{r}}$ changes orientation from one sub-aperture to another.

4.2. Numerical results. The setup for the numerical results is the same as the one used in section 3.4 to obtain Figures 3.3–3.4. The data are generated using the single scattering model (4.1), for a reflectivity function that changes with the direction of illumination and is supported at two points at distance of order $\overline{\lambda}\,\overline{L}_o/A$, where $\overline{L}_o = |\overline{\boldsymbol{r}}_o - \overline{\boldsymbol{y}}|$.

We display in Figure 4.2 the six entries of the row vectors $\boldsymbol{x}_{q\rightarrow}$ defined in (4.10), as $\overline{\boldsymbol{r}}$ varies in the large aperture, and for points in Ω indexed by q, separated by distances $\overline{\lambda}\,\overline{L}_o/A$ in cross-range. The plots in the bottom line of Figure 4.2 show that the MMV method gives a good estimate of these row vectors, for noiseless data.

In Figure 4.3 we display the components $(\boldsymbol{v}_j^T \boldsymbol{\rho}_q \boldsymbol{v}_l)_{j,l=1,2}$ of the reflectivity matrix $\boldsymbol{\rho}_q$ and its reconstruction, for each sub-aperture centered at $\overline{\boldsymbol{r}}$. As in the note at the end of the previous section, we let $\{\boldsymbol{v}_j\}_{j=1,2,3}$ be the orthonormal basis of eigenvectors of the approximation (4.8) of the Green's tensor, with $\boldsymbol{v}_3$ along $\overline{\boldsymbol{r}}$. The reconstruction displayed in Figure 4.3 is calculated as follows: With the estimated vectors $\boldsymbol{x}_{q\rightarrow}$ displayed in Figure 4.2 we calculate the minimum ℓ_2 norm solution of (4.10), using the truncated SVD of the singular matrix $\boldsymbol{\Gamma}(\overline{\boldsymbol{r}})$. This corresponds to setting to zero the components $\boldsymbol{v}_j^T \boldsymbol{\rho}_q \boldsymbol{v}_l$ of the estimated $\boldsymbol{\rho}_q$, for either j or l equal to 3. The other components are displayed in the figure, and they are well reconstructed.

5. Proofs. Here we prove the results stated in section 2.2. We begin with a Lemma, in section 5.1, which we then use in sections 5.2 and 5.3 to prove Theorems 2.1 and 2.2. Proposition 2.3 is proved in section 5.4 and the results for the clusters are proved in section 5.5.

5.1. A basic lemma. Let us denote by $\widehat{\boldsymbol{X}}$ the matrix obtained by normalizing the nonzero rows in $\boldsymbol{X}$, the unknown in the inverse problem,

$$\widehat{\boldsymbol{x}}_{q\rightarrow} = \frac{\boldsymbol{x}_{q\rightarrow}}{\|\boldsymbol{x}_{q\rightarrow}\|_2}, \quad q \in \mathcal{S}, \tag{5.1}$$

19

where $\| \cdot \|$ is the norm induced by the usual Hermitian inner product. Introduce the linear operator $\mathfrak{L} : \mathbb{C}^{N_r \times N_v} \to \mathbb{C}$ defined by

$$\mathfrak{L}(\boldsymbol{V}) = \mathrm{tr}\Big[(\boldsymbol{G}\widehat{\boldsymbol{X}})^\star \boldsymbol{V}\Big], \quad \forall\, \boldsymbol{V} \in \mathbb{C}^{N_r \times N_v}, \tag{5.2}$$

where $\mathrm{tr}[\cdot]$ denotes the trace. We have the following result:

LEMMA 5.1. *The linear operator $\mathfrak{L}$ defined in (5.2) satisfies the inequality*

$$\big|\mathfrak{L}(\boldsymbol{V})\big| \leq \|(\boldsymbol{G}^\star \boldsymbol{V})_{\mathcal{S}\to}\|_{1,2}, \tag{5.3}$$

for any $\boldsymbol{V} \in \mathbb{C}^{N_r \times N_v}$. The matrix $\boldsymbol{X}$ satisfies the inequality

$$\|\boldsymbol{X}\|_{1,2}\big(1 - \mathscr{I}_{N_v}\big) \leq \big|\mathfrak{L}(\boldsymbol{GX})\big|, \tag{5.4}$$

and with r, $\boldsymbol{X}^{\varepsilon,r}$ and $\boldsymbol{E}^{\varepsilon,r}$ defined as in Theorem 2.1, we have

$$\big|\mathfrak{L}(\boldsymbol{GX}^{\varepsilon,r})\big| \leq \big(1 + \mathscr{I}_{N_v}\big)\|\boldsymbol{X}^{\varepsilon,r}\|_{1,2}, \tag{5.5}$$
$$\big|\mathfrak{L}(\boldsymbol{GE}^{\varepsilon,r})\big| \leq \big(1 - r + \mathscr{I}_{N_v}\big)\|\boldsymbol{E}^{\varepsilon,r}\|_{1,2}. \tag{5.6}$$

Proof. We start with definition (5.2), and use the invariance of the trace under cyclic permutations, and the row support $\mathcal{S}$ of $\boldsymbol{X}$, to obtain

$$\begin{aligned}
\mathfrak{L}(\boldsymbol{GX}) &= \mathrm{tr}\Big[(\boldsymbol{G}\widehat{\boldsymbol{X}})^\star \boldsymbol{GX}\Big] \\
&= \mathrm{tr}\Big[\widehat{\boldsymbol{X}}^\star \boldsymbol{G}^\star \boldsymbol{GX}\Big] \\
&= \mathrm{tr}\Big[\boldsymbol{X}\widehat{\boldsymbol{X}}^\star \boldsymbol{G}^\star \boldsymbol{G}\Big] \\
&= \sum_{j,q\in\mathcal{S}} (\boldsymbol{X}\widehat{\boldsymbol{X}}^\star)_{j,q}(\boldsymbol{G}^\star \boldsymbol{G})_{q,j} \\
&= \sum_{j,q\in\mathcal{S}} \langle \boldsymbol{x}_{j\to}, \widehat{\boldsymbol{x}}_{q\to}\rangle \langle \boldsymbol{g}_q, \boldsymbol{g}_j\rangle.
\end{aligned}$$

We rewrite this further with the normalization condition (2.15) and definition (2.13), and use the triangle inequality to obtain the bound

$$\begin{aligned}
\big|\mathfrak{L}(\boldsymbol{GX})\big| &= \Big|\sum_{q\in\mathcal{S}} \Big[\langle \boldsymbol{x}_{q\to}, \widehat{\boldsymbol{x}}_{q\to}\rangle \langle \boldsymbol{g}_q, \boldsymbol{g}_q\rangle + \sum_{j\in\mathcal{S}\backslash\{q\}} \langle \boldsymbol{x}_{j\to}, \widehat{\boldsymbol{x}}_{q\to}\rangle \langle \boldsymbol{g}_q, \boldsymbol{g}_j\rangle\Big]\Big| \\
&= \Big|\sum_{q\in\mathcal{S}} \|\boldsymbol{x}_{q\to}\|_2\Big[1 + \sum_{j\in\mathcal{S}\backslash\{q\}} \mu(\boldsymbol{x}_{j\to}, \boldsymbol{x}_{q\to})\mu(\boldsymbol{g}_q, \boldsymbol{g}_j)\Big]\Big| \\
&\geq \sum_{q\in\mathcal{S}} \|\boldsymbol{x}_{q\to}\|_2\Big[1 - \sum_{j\in\mathcal{S}\backslash\{q\}} |\mu(\boldsymbol{x}_{j\to}, \boldsymbol{x}_{q\to})||\mu(\boldsymbol{g}_q, \boldsymbol{g}_j)|\Big]\Big| \\
&\geq \sum_{q\in\mathcal{S}} \|\boldsymbol{x}_{q\to}\|_2(1 - \mathscr{I}_{N_v}).
\end{aligned}$$

The result (5.4) follows from definition (2.9) of the matrix norm $\|\boldsymbol{X}\|_{1,2}$.

Similarly,

$$\mathfrak{L}(\boldsymbol{GX}^{\varepsilon,r}) = \sum_{j\in\mathcal{S}^\varepsilon} \sum_{q\in\mathcal{S}} \langle \boldsymbol{x}^{\varepsilon,r}_{j\to}, \widehat{\boldsymbol{x}}_{q\to}\rangle \langle \boldsymbol{g}_q, \boldsymbol{g}_j\rangle = \sum_{j\in\mathcal{S}^\varepsilon} \|\boldsymbol{x}^{\varepsilon,r}_{j\to}\|_2 \sum_{q\in\mathcal{S}} \mu(\boldsymbol{x}^{\varepsilon,r}_{j\to}, \widehat{\boldsymbol{x}}_{q\to})\mu(\boldsymbol{g}_q, \boldsymbol{g}_j),$$

where $\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r}$ denotes the j–th row of $\boldsymbol{X}^{\varepsilon,r}$. Using the decomposition (2.26) of the row support $\mathcal{S}^\varepsilon$ of $\boldsymbol{X}^{\varepsilon,r}$,

$$
\begin{aligned}
\big|\mathfrak{L}(\boldsymbol{G}\boldsymbol{X}^{\varepsilon,r})\big| &= \Big|\sum_{i\in\mathcal{S}}\sum_{j\in\mathfrak{S}_i}\|\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r}\|_2\sum_{q\in\mathcal{S}}\mu(\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{q\rightarrow})\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)\Big| \\
&= \Big|\sum_{i\in\mathcal{S}}\sum_{j\in\mathfrak{S}_i}\|\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r}\|_2\Big[\mu(\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{i\rightarrow})\mu(\boldsymbol{g}_i,\boldsymbol{g}_j)+\sum_{q\in\mathcal{S}\setminus\{i\}}\mu(\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{q\rightarrow})\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)\Big]\Big|.
\end{aligned}
$$

By the construction in (2.26), for any $j\in\mathfrak{S}_i$, the index $n(j)\in\mathcal{S}$ of the nearest point to $\boldsymbol{y}_j$ is $n(j)=i$, so the sum in q is over the set $\mathcal{S}\setminus\{n(j)\}$. Using the triangle inequality and the definition (2.13) of $\mathscr{I}_{N_v}$, we get

$$
\begin{aligned}
\big|\mathfrak{L}(\boldsymbol{G}\boldsymbol{X}^{\varepsilon,r})\big| &\leq \sum_{i\in\mathcal{S}}\sum_{j\in\mathfrak{S}_i}\|\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r}\|_2\Big[|\mu(\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{i\rightarrow})\mu(\boldsymbol{g}_i,\boldsymbol{g}_j)|+\sum_{q\in\mathcal{S}\setminus\{n(j)\}}|\mu(\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{q\rightarrow})|\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)|\Big] \\
&\leq \sum_{i\in\mathcal{S}}\sum_{j\in\mathfrak{S}_i}\|\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r}\|_2(1+\mathscr{I}_{N_v}) \\
&= \sum_{j\in\mathcal{S}^\varepsilon}\|\boldsymbol{x}_{j\rightarrow}^{\varepsilon,r}\|_2(1+\mathscr{I}_{N_v}).
\end{aligned}
$$

Since $\mathcal{S}^\varepsilon$ is the row support of $\boldsymbol{X}^{\varepsilon,r}$, we can extend the sum to $1\leq j\leq N_{\boldsymbol{y}}$, and the result (5.5) follows from the definition of the $\|\cdot\|_{1,2}$ norm.

To prove (5.6), recall that $\boldsymbol{E}^{\varepsilon,r}$ is supported by definition outside the r–vicinity of $\mathcal{S}$, in the set

$$
\mathfrak{B}_r^c(\mathcal{S})=\{1,\ldots,N_{\boldsymbol{y}}\}\setminus\mathfrak{B}_r(\mathcal{S}).
$$

Then, if we denote by $\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r}$ the rows of $\boldsymbol{E}^{\varepsilon,r}$, we have

$$
\begin{aligned}
\mathfrak{L}(\boldsymbol{G}\boldsymbol{E}^{\varepsilon,r}) &= \sum_{j\in\mathfrak{B}_r^c(\mathcal{S})}\sum_{q\in\mathcal{S}}\big\langle\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{q\rightarrow}\big\rangle\langle\boldsymbol{g}_q,\boldsymbol{g}_j\rangle \\
&= \sum_{j\in\mathfrak{B}_r^c(\mathcal{S})}\|\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r}\|_2\sum_{q\in\mathcal{S}}\mu(\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{q\rightarrow})\mu(\boldsymbol{g}_q,\boldsymbol{g}_j) \\
&= \sum_{j\in\mathfrak{B}_r^c(\mathcal{S})}\|\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r}\|_2\Big[\mu(\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{n(j)\rightarrow})\mu(\boldsymbol{g}_{n(j)},\boldsymbol{g}_j)+\sum_{q\in\mathcal{S}\setminus\{n(j)\}}\mu(\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{q\rightarrow})\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)\Big].
\end{aligned}
$$

Taking the absolute value and using the triangle inequality and definition (2.13) of $\mathscr{I}_{N_v}$, we obtain the bound

$$
\big|\mathfrak{L}(\boldsymbol{G}\boldsymbol{E}^{\varepsilon,r})\big|\leq\sum_{j\in\mathfrak{B}_r^c(\mathcal{S})}\|\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r}\|_2\Big[|\mu(\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{n(j)\rightarrow})||\mu(\boldsymbol{g}_{n(j)},\boldsymbol{g}_j)|+\mathscr{I}_{N_v}\Big].
$$

But $|\mu(\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r},\widehat{\boldsymbol{x}}_{n(j)\rightarrow})|\leq 1$ and $|\mu(\boldsymbol{g}_{n(j)},\boldsymbol{g}_j)|=1-\mathfrak{D}(j,n(j))$, with $\mathfrak{D}(j,n(j))\geq r$ for any $j\in\mathfrak{B}_r^c(\mathcal{S})$, so the bound becomes

$$
\big|\mathfrak{L}(\boldsymbol{G}\boldsymbol{E}^{\varepsilon,r})\big|\leq\sum_{j\in\mathfrak{B}_r^c(\mathcal{S})}\|\boldsymbol{e}_{j\rightarrow}^{\varepsilon,r}\|_2(1-r+\mathscr{I}_{N_v}).
$$

We can extend the sum to $1\leq j\leq N_{\boldsymbol{y}}$ because $\boldsymbol{E}^{\varepsilon,r}$ is supported in $\mathfrak{B}_r^c(\mathcal{S})$, and the result (5.6) follows from the definition of the $\|\cdot\|_{1,2}$ norm.

21

Finally, for any $\boldsymbol{V} \in \mathbb{C}^{N_r \times N_v}$, we obtain using the invariance of the trace to cyclic permutations that

$$
\begin{aligned}
\mathfrak{L}(\boldsymbol{V}) &= \operatorname{tr}\Big[(\boldsymbol{G}\widehat{\boldsymbol{X}})^{\star}\boldsymbol{V}\Big] \\
&= \operatorname{tr}\Big[\widehat{\boldsymbol{X}}^{\star}\boldsymbol{G}^{\star}\boldsymbol{V}\Big] \\
&= \operatorname{tr}\Big[\boldsymbol{G}^{\star}\boldsymbol{V}\widehat{\boldsymbol{X}}^{\star}\Big] \\
&= \sum_{j=1}^{N_y}\big\langle(\boldsymbol{G}^{\star}\boldsymbol{V})_{j\rightarrow},\widehat{\boldsymbol{x}}_{j\rightarrow}\big\rangle \\
&= \sum_{j\in\mathcal{S}}\big\langle(\boldsymbol{G}^{\star}\boldsymbol{V})_{j\rightarrow},\widehat{\boldsymbol{x}}_{j\rightarrow}\big\rangle,
\end{aligned}
$$

where the last equality is because $\boldsymbol{X}$ is row supported in $\mathcal{S}$. Taking the absolute value and using the triangle and Cauchy-Schwartz inequalities we get

$$
\begin{aligned}
\big|\mathfrak{L}(\boldsymbol{V})\big| &\leq \sum_{j\in\mathcal{S}}\big|\big\langle(\boldsymbol{G}^{\star}\boldsymbol{V})_{j\rightarrow},\widehat{\boldsymbol{x}}_{j\rightarrow}\big\rangle\big| \\
&\leq \sum_{j\in\mathcal{S}}\|(\boldsymbol{G}^{\star}\boldsymbol{V})_{j\rightarrow}\|_2 \\
&= \|(\boldsymbol{G}^{\star}\boldsymbol{V})_{\mathcal{S}\rightarrow}\|_{1,2}.
\end{aligned}
$$

This is the result (5.3) in the lemma. □

5.2. Proof of Theorem 2.1. The bound (2.19) follows from the definition of $\boldsymbol{W}^{\varepsilon}$ and the triangle inequality,

$$
\|\boldsymbol{W}^{\varepsilon}\|_F = \|\boldsymbol{D}_{\boldsymbol{W}} - \boldsymbol{G}\boldsymbol{X}^{\varepsilon} - \boldsymbol{W}\|_F \leq \|\boldsymbol{D}_{\boldsymbol{W}} - \boldsymbol{G}\boldsymbol{X}^{\varepsilon}\|_F + \|\boldsymbol{W}\|_F \leq 2\varepsilon,
$$

where we used the assumption (2.12) and that $\boldsymbol{X}^{\varepsilon}$ is the minimizer of (2.11).

Using again the definition of $\boldsymbol{W}^{\varepsilon}$ and the linearity of the operator (5.2), we write

$$
\mathfrak{L}(\boldsymbol{G}\boldsymbol{X}) + \mathfrak{L}(\boldsymbol{W}^{\varepsilon}) = \mathfrak{L}(\boldsymbol{G}\boldsymbol{X} + \boldsymbol{W}^{\varepsilon}) = \mathfrak{L}(\boldsymbol{G}\boldsymbol{X}^{\varepsilon}) = \mathfrak{L}(\boldsymbol{G}\boldsymbol{X}^{\varepsilon,r}) + \mathfrak{L}(\boldsymbol{G}\boldsymbol{E}^{\varepsilon,r}),
$$

where the last equality is by the decomposition (2.20). The result (5.4) in Lemma 5.1 gives

$$
\|\boldsymbol{X}\|_{1,2}(1 - \mathscr{I}_{N_v}) \leq \big|\mathfrak{L}(\boldsymbol{G}\boldsymbol{X})\big| = \big|\mathfrak{L}(\boldsymbol{G}\boldsymbol{X}^{\varepsilon,r}) + \mathfrak{L}(\boldsymbol{G}\boldsymbol{E}^{\varepsilon,r}) - \mathfrak{L}(\boldsymbol{W}^{\varepsilon})\big|,
$$

and using the triangle inequality and the estimates (5.3), (5.5) and (5.6), we get

$$
\begin{aligned}
\|\boldsymbol{X}\|_{1,2}(1 - \mathscr{I}_{N_v}) &\leq \big|\mathfrak{L}(\boldsymbol{G}\boldsymbol{X}^{\varepsilon,r})\big| + \big|\mathfrak{L}(\boldsymbol{G}\boldsymbol{E}^{\varepsilon,r})\big| + \big|\mathfrak{L}(\boldsymbol{W}^{\varepsilon})\big| \\
&\leq (1 + \mathscr{I}_{N_v})\|\boldsymbol{X}^{\varepsilon,r}\|_{1,2} + (1 - r + \mathscr{I}_{N_v})\|\boldsymbol{E}^{\varepsilon,r}\|_{1,2} + \|(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon})_{\mathcal{S}\rightarrow}\|_{1,2}. \quad (5.7)
\end{aligned}
$$

Note that by (2.10) and (2.12),

$$
\|\boldsymbol{G}\boldsymbol{X} - \boldsymbol{D}_{\boldsymbol{W}}\|_F = \|\boldsymbol{W}\|_F < \varepsilon,
$$

so since $\boldsymbol{X}^{\varepsilon}$ is the minimizer of (2.11), we must have

$$
\|\boldsymbol{X}^{\varepsilon}\|_{1,2} \leq \|\boldsymbol{X}\|_{1,2}. \quad (5.8)
$$

We also have from the decomposition (2.20) of $\boldsymbol{X}^{\varepsilon}$ in the matrices $\boldsymbol{X}^{\varepsilon,r}$ and $\boldsymbol{E}^{\varepsilon,r}$ with disjoint row support that

$$
\|\boldsymbol{X}^{\varepsilon}\|_{1,2} = \|\boldsymbol{X}^{\varepsilon,r} + \boldsymbol{E}^{\varepsilon,r}\|_{1,2} = \|\boldsymbol{X}^{\varepsilon,r}\|_{1,2} + \|\boldsymbol{E}^{\varepsilon,r}\|_{1,2}. \quad (5.9)
$$

Substituting in (5.7) we get

$$
\begin{aligned}
\|\boldsymbol{X}^{\varepsilon}\|_{1,2}(1 - \mathscr{I}_{N_v}) &\leq (1 + \mathscr{I}_{N_v})(\|\boldsymbol{X}^{\varepsilon}\|_{1,2} - \|\boldsymbol{E}^{\varepsilon,r}\|_{1,2}) + (1 - r + \mathscr{I}_{N_v})\|\boldsymbol{E}^{\varepsilon,r}\|_{1,2} + \|(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon})_{\mathcal{S}\rightarrow}\|_{1,2} \\
&= (1 + \mathscr{I}_{N_v})\|\boldsymbol{X}^{\varepsilon}\|_{1,2} - r\|\boldsymbol{E}^{\varepsilon,r}\|_{1,2} + \|(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon})_{\mathcal{S}\rightarrow}\|_{1,2},
\end{aligned}
$$

and the result (2.21) stated in the theorem follows. □

5.3. Proof of Theorem 2.2. Let us start with the definition of the matrices $\boldsymbol{W}^{\varepsilon}$, $\boldsymbol{X}^{\varepsilon,r}$ and $\boldsymbol{E}^{\varepsilon,r}$ given in Theorem 2.1, and write

$$\boldsymbol{G}\boldsymbol{X}^{\varepsilon} = \boldsymbol{G}(\boldsymbol{X}^{\varepsilon,r} + \boldsymbol{E}^{\varepsilon,r}) = \boldsymbol{G}\boldsymbol{X} + \boldsymbol{W}^{\varepsilon}.$$

With the decomposition (2.23) of $\boldsymbol{X}^{\varepsilon,r}$, we get

$$\boldsymbol{G}(\boldsymbol{X} - \boldsymbol{\mathfrak{X}}^{\varepsilon,r}) = \boldsymbol{G}\boldsymbol{E}^{\varepsilon,r} + \boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r} - \boldsymbol{W}^{\varepsilon}, \tag{5.10}$$

and we prove next the analogue of the result (5.7) for $\boldsymbol{X}$ replaced by the matrix $\boldsymbol{X} - \boldsymbol{\mathfrak{X}}^{\varepsilon,r}$ and $\boldsymbol{X}^{\varepsilon,r}$ replaced by 0. Looking at the proof of (5.4) in section 5.1, we note that we only used that $\boldsymbol{X}$ has row support in $\mathcal{S}$. The same holds for the matrix $\boldsymbol{X} - \boldsymbol{\mathfrak{X}}^{\varepsilon,r}$, so we can write directly the analogue of (5.4)

$$\|\boldsymbol{X} - \boldsymbol{\mathfrak{X}}^{\varepsilon,r}\|_{1,2}(1 - \mathscr{I}_{N_v}) \leq \left| \mathfrak{L}\big(\boldsymbol{G}(\boldsymbol{X} - \boldsymbol{\mathfrak{X}}^{\varepsilon,r})\big) \right|. \tag{5.11}$$

The right hand side in this equation can be estimated using (5.10) and the linearity of the operator $\mathfrak{L}$,

$$\left| \mathfrak{L}\big(\boldsymbol{G}(\boldsymbol{X} - \boldsymbol{\mathfrak{X}}^{\varepsilon,r})\big) \right| = \left| \mathfrak{L}(\boldsymbol{G}\boldsymbol{E}^{\varepsilon,r}) + \mathfrak{L}(\boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r} - \boldsymbol{W}^{\varepsilon}) \right| \leq \left| \mathfrak{L}(\boldsymbol{G}\boldsymbol{E}^{\varepsilon,r}) \right| + \left| \mathfrak{L}(\boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r} - \boldsymbol{W}^{\varepsilon}) \right|.$$

Substituting in (5.11) and using the estimates (5.6) and (5.3), with $\boldsymbol{V}$ replaced by $\boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r} - \boldsymbol{W}^{\varepsilon}$, we obtain

$$\|\boldsymbol{X} - \boldsymbol{\mathfrak{X}}^{\varepsilon,r}\|_{1,2}(1 - \mathscr{I}_{N_v}) \leq (1 - r + \mathscr{I}_{N_v})\|\boldsymbol{E}^{\varepsilon,r}\|_{1,2} + \left\| \big(\boldsymbol{G}^{\star}(\boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r} - \boldsymbol{W}^{\varepsilon})\big)_{\mathcal{S}\to} \right\|_{1,2}.$$

But, by equation (2.25),

$$\big(\boldsymbol{G}^{\star}\boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r}\big)_{\mathcal{S}\to} = \boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}\boldsymbol{\mathcal{E}}^{\varepsilon,r} = 0,$$

and the desired estimate is

$$\|\boldsymbol{X} - \boldsymbol{\mathfrak{X}}^{\varepsilon,r}\|_{1,2}(1 - \mathscr{I}_{N_v}) \leq (1 - r + \mathscr{I}_{N_v})\|\boldsymbol{E}^{\varepsilon,r}\|_{1,2} + \left\| \big(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon}\big)_{\mathcal{S}\to} \right\|_{1,2}. \tag{5.12}$$

Next, we substitute the bound (2.21) on the error term $\boldsymbol{E}^{\varepsilon,r}$ in (5.12), and obtain after simple algebraic manipulations that

$$\|\boldsymbol{X} - \boldsymbol{\mathfrak{X}}^{\varepsilon,r}\|_{1,2} \leq \frac{2\mathscr{I}_{N_v}(1 - r + \mathscr{I}_{N_v})}{r(1 - \mathscr{I}_{N_v})}\|\boldsymbol{X}^{\varepsilon}\|_{1,2} + \frac{(1 + \mathscr{I}_{N_v})}{r(1 - \mathscr{I}_{N_v})}\left\| \big(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon}\big)_{\mathcal{S}\to} \right\|_{1,2}. \tag{5.13}$$

The assumption $2\mathscr{I}_{N_v} < r < 1$ implies that

$$1 - r + \mathscr{I}_{N_v} \leq 1 - 2\mathscr{I}_{N_v} + \mathscr{I}_{N_v} = 1 - \mathscr{I}_{N_v},$$

and moreover

$$\frac{1 + \mathscr{I}_{N_v}}{1 - \mathscr{I}_{N_v}} < \frac{1 + \mathscr{I}_{N_v}}{1 - r/2} < 2(1 + \mathscr{I}_{N_v}) < 3.$$

Substituting in (5.13) we obtain the result (2.28) of Theorem 2.2.

It remains to prove the estimate (2.29). We begin with the identity

$$\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}} = \boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}} - \boldsymbol{\mathcal{E}}^{\varepsilon,r},$$

and use equation (2.25) to conclude that

$$\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}(\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}}) = \boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}}).$$

23

By construction, both $\boldsymbol{\mathfrak{X}}^{\varepsilon,r}$ and $\overline{\boldsymbol{X}^{\varepsilon,r}}$ are row supported in $\mathcal{S}$, so we can rewrite this equation as

$$(\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{\mathcal{S}\to} - (\boldsymbol{I} - \boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}_{\mathcal{S}})(\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{\mathcal{S}\to} = \boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}}), \tag{5.14}$$

where $\boldsymbol{I}$ is the $|\mathcal{S}| \times |\mathcal{S}|$ identity matrix. We now estimate each term in this equation.

For the right hand side in (5.14) we have

$$\begin{aligned}
\|\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})\|_{1,1} &= \sum_{q\in\mathcal{S}}\sum_{v=1}^{N_v}\left|\left(\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})\right)_{q,v}\right| \\
&= \sum_{q\in\mathcal{S}}\sum_{v=1}^{N_v}\left|\sum_{j=1}^{N_{\boldsymbol{y}}}(\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G})_{q,j}(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v}\right| \\
&= \sum_{q\in\mathcal{S}}\sum_{v=1}^{N_v}\left|\sum_{j\in\mathcal{S}\cup\mathcal{S}^{\varepsilon}}\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v}\right| \\
&= \sum_{q\in\mathcal{S}}\sum_{v=1}^{N_v}\left|\sum_{j\in\mathcal{S}\cup\mathcal{S}^{\varepsilon}\setminus\mathfrak{S}_q}\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v} + \sum_{j\in(\mathcal{S}\cup\mathcal{S}^{\varepsilon})\cap\mathfrak{S}_q}\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v}\right|, \tag{5.15}
\end{aligned}$$

where the first two equalities are by the definition of the norm and of the matrix product, and the third equality uses the definition (2.14) and the row support $\mathcal{S}\cup\mathcal{S}^{\varepsilon}$ of $\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}}$. Now let us recall the definition (2.27) of $\overline{\boldsymbol{X}^{\varepsilon,r}}$, and the decomposition (2.26) of the support $\mathcal{S}^{\varepsilon}$ of $\boldsymbol{X}^{\varepsilon,r}$, to obtain

$$\sum_{j\in(\mathcal{S}\cup\mathcal{S}^{\varepsilon})\cap\mathfrak{S}_q}\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)\boldsymbol{X}_{j,v}^{\varepsilon,r} = \sum_{j\in\mathfrak{S}_q}\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)\boldsymbol{X}_{j,v}^{\varepsilon,r} = \overline{\boldsymbol{X}_{q,v}^{\varepsilon,r}},$$

and

$$\overline{\boldsymbol{X}_{j,v}^{\varepsilon,r}} = \overline{\boldsymbol{X}_{q,v}^{\varepsilon,r}}\,\delta_{j,q}, \quad \forall\, j\in\mathfrak{S}_q,$$

where $\delta_{j,q}$ is the Kronecker delta symbol. Since $\mu(\boldsymbol{g}_q,\boldsymbol{g}_q) = 1$, we conclude that the second term in (5.15) vanishes and the result becomes

$$\begin{aligned}
\|\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})\|_{1,1} &= \sum_{q\in\mathcal{S}}\sum_{v=1}^{N_v}\left|\sum_{j\in\mathcal{S}\cup\mathcal{S}^{\varepsilon}\setminus\mathfrak{S}_q}\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v}\right| \\
&\leq \sum_{v=1}^{N_v}\sum_{q\in\mathcal{S}}\sum_{j\in\mathcal{S}\cup\mathcal{S}^{\varepsilon}\setminus\mathfrak{S}_q}|\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)|\left|(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v}\right|. \tag{5.16}
\end{aligned}$$

Note that the set $\{(j,q) : j\in\mathcal{S}\cup\mathcal{S}^{\varepsilon}\setminus\mathfrak{S}_q,\ q\in\mathcal{S}\}$ is the same as the set $\{(j,q) : j\in\mathcal{S}\cup\mathcal{S}^{\varepsilon},\ q\in\mathcal{S}\setminus\{n(j)\}\}$, so we can rewrite (5.16) as

$$\|\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})\|_{1,1} \leq \sum_{v=1}^{N_v}\sum_{j\in\mathcal{S}\cup\mathcal{S}^{\varepsilon}}\left|(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v}\right|\sum_{q\in\mathcal{S}\setminus\{n(j)\}}|\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)|.$$

The last sum in this equation is bounded above by the interaction coefficient $\mathscr{I}_1$, and using the definition of the $\|\cdot\|_{1,1}$ norm we get

$$\|\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})\|_{1,1} \leq \mathscr{I}_1\|\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}}\|_{1,1}. \tag{5.17}$$

24

With a similar calculation we obtain

$$
\begin{aligned}
\left\|(\boldsymbol{I} - \boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}_{\mathcal{S}})(\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})\right\|_{1,1} &= \sum_{q\in\mathcal{S}}\sum_{v=1}^{N_v}\left|\sum_{j\in\mathcal{S}}(\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}_{\mathcal{S}} - \boldsymbol{I})_{q,j}(\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v}\right| \\
&\leq \sum_{j\in\mathcal{S}}\sum_{v=1}^{N_v}|(\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v}|\sum_{q\in\mathcal{S}}|\mu(\boldsymbol{g}_q,\boldsymbol{g}_j) - \delta_{q,j}| \\
&= \sum_{j\in\mathcal{S}}\sum_{v=1}^{N_v}|(\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})_{j,v}|\sum_{q\in\mathcal{S}\setminus\{j\}}|\mu(\boldsymbol{g}_q,\boldsymbol{g}_j)|,
\end{aligned}
$$

where we used the triangle inequality, the identity $(\boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}_{\mathcal{S}})_{q,j} = \mu(\boldsymbol{g}_q,\boldsymbol{g}_j)$ and $\mu(\boldsymbol{g}_q,\boldsymbol{g}_q) = 1$. The last sum is bounded above by the interaction coefficient $\mathscr{I}_1$, and using that $\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}}$ is row supported in $\mathcal{S}$, and the definition of the $\|\cdot\|_{1,1}$ norm, we get

$$
\left\|(\boldsymbol{I} - \boldsymbol{G}_{\mathcal{S}}^{\star}\boldsymbol{G}_{\mathcal{S}})(\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}})\right\|_{1,1} \leq \mathscr{I}_1\|\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}}\|_{1,1}. \tag{5.18}
$$

Gathering the results (5.14), (5.17)–(5.18), and using the triangle inequality, we obtain the bound

$$
(1 - \mathscr{I}_1)\|\boldsymbol{\mathfrak{X}}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}}\|_{1,1} \leq \mathscr{I}_1\|\boldsymbol{X}^{\varepsilon,r} - \overline{\boldsymbol{X}^{\varepsilon,r}}\|_{1,1} \leq \mathscr{I}_1\left(\|\boldsymbol{X}^{\varepsilon,r}\|_{1,1} + \|\overline{\boldsymbol{X}^{\varepsilon,r}}\|_{1,1}\right). \tag{5.19}
$$

We also have from the definition (2.27) and the inequality $|\mu(\boldsymbol{g}_j,\boldsymbol{g}_l)| \leq 1$ for all $j,l = 1,\ldots,N_{\boldsymbol{y}}$, that

$$
\|\overline{\boldsymbol{X}^{\varepsilon,r}}\|_{1,1} \leq \|\boldsymbol{X}^{\varepsilon,r}\|_{1,1}.
$$

The estimate (2.29) in Theorem 2.2 follows by substituting this in (5.19). $\square$

5.4. Proof of Proposition 2.3. Recall the definition (5.1) of the unit row vectors $\widehat{\boldsymbol{x}}_{q\rightarrow}$. Because the rows of $\boldsymbol{X}$ are assumed orthogonal in the proposition, $\{\widehat{\boldsymbol{x}}_{q\rightarrow}, q \in \mathcal{S}\}$ is an orthonormal subset of $\mathbb{C}^{1\times N_v}$, and we conclude from Bessel's inequality that

$$
\sum_{q\in\mathcal{S}\setminus\{n(j)\}}|\langle\boldsymbol{v}_{\rightarrow},\widehat{\boldsymbol{x}}_{q\rightarrow}\rangle|^2 \leq \|\boldsymbol{v}_{\rightarrow}\|_2^2, \quad \forall\,\boldsymbol{v}_{\rightarrow} \in \mathbb{C}^{1\times N_v} \text{ and } j = 1,\ldots,N_{\boldsymbol{y}}.
$$

Dividing both sides in this equation by $\|\boldsymbol{v}_{\rightarrow}\|_2^2$ and recalling definition (2.18), we obtain

$$
\sum_{q\in\mathcal{S}\setminus\{n(j)\}}|\mu(\boldsymbol{v}_{\rightarrow},\widehat{\boldsymbol{x}}_{q\rightarrow})|^2 \leq 1, \quad \forall\,\boldsymbol{v}_{\rightarrow} \in \mathbb{C}^{1\times N_v} \text{ and } j = 1,\ldots,N_{\boldsymbol{y}}. \tag{5.20}
$$

For a given j and v, we define the vector $\boldsymbol{\nu}^{(j,\boldsymbol{v}_{\rightarrow})} \in \mathbb{R}^{1\times(|\mathcal{S}|-1)}$ with entries $|\mu(\boldsymbol{v}_{\rightarrow},\widehat{\boldsymbol{x}}_{q\rightarrow})|$. Recall also from section 2.2.3 the vector $\boldsymbol{\gamma}^{(j)} \in \mathbb{R}^{1\times(|\mathcal{S}|-1)}$ with entries $|\mu(\boldsymbol{g}_j,\boldsymbol{g}_q)|$, for $q \in \mathcal{S}\setminus\{n(j)\}$, which is a set with cardinality $|\mathcal{S}| - 1$. Using these vectors, we have

$$
\sup_{\boldsymbol{v}_{\rightarrow}\in\mathbb{C}^{1\times N_v}}\sum_{q\in\mathcal{S}\setminus\{n(j)\}}|\mu(\boldsymbol{g}_j,\boldsymbol{g}_q)||\mu(\boldsymbol{v}_{\rightarrow},\widehat{\boldsymbol{x}}_{q\rightarrow})| = \sup_{\boldsymbol{\nu}^{(j,\boldsymbol{v}_{\rightarrow})}\in\mathbb{R}^{1\times|\mathcal{S}|-1},\|\boldsymbol{\nu}^{(j,\boldsymbol{v}_{\rightarrow})}\|\leq 1}\left(\boldsymbol{\nu}^{(j,\boldsymbol{v}_{\rightarrow})},\boldsymbol{\gamma}^{(j)}\right) = \|\boldsymbol{\gamma}^{(j)}\|_2,
$$

where $(\cdot,\cdot)$ denotes the Euclidian inner product in $\mathbb{R}^{1\times|\mathcal{S}|-1}$ and we used inequality (5.20) to conclude that $\boldsymbol{\nu}^{(j)}$ must lie in the unit ball in $\mathbb{R}^{1\times|\mathcal{S}|-1}$. The last equality is because the sup is achieved for

$$
\boldsymbol{\nu}^{(j,\boldsymbol{v}_{\rightarrow})} = \boldsymbol{\gamma}^{(j)}/\|\boldsymbol{\gamma}^{(j)}\|_2.
$$

Substituting in the definition (2.13), we obtain the result (2.30). $\square$

5.5. Proof of cluster results. The proof of Theorem 2.5 is the same as in section 5.2, with $\boldsymbol{X}$ replaced by $\boldsymbol{U}$, $\mathcal{S}$ replaced by $\mathcal{C}$ and $\boldsymbol{W}$ replaced by $\boldsymbol{\mathcal{W}}$. This leads to the estimate

$$\|\boldsymbol{E}^{\varepsilon,r}\|_{1,2} \leq \frac{2\mathscr{I}_{N_v}^{\boldsymbol{U}}}{r}\|\boldsymbol{X}^{\varepsilon}\|_{1,2} + \frac{1}{r}\left\|\left(\boldsymbol{G}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon} - \boldsymbol{X} + \boldsymbol{R})\right)_{\mathcal{C}\rightarrow}\right\|_{1,2}, \tag{5.21}$$

where we used that $\boldsymbol{U} = \boldsymbol{X} - \boldsymbol{R}$ in the last term. But equation (2.38) implies that

$$\boldsymbol{G}_{\mathcal{C}}^{\dagger}\boldsymbol{G}\boldsymbol{R} = \boldsymbol{G}_{\mathcal{C}}^{\dagger}\boldsymbol{G}(\boldsymbol{X} - \boldsymbol{U}) = \boldsymbol{G}_{\mathcal{C}}^{\dagger}\boldsymbol{G}\boldsymbol{X} - \boldsymbol{G}_{\mathcal{C}}^{\dagger}\boldsymbol{G}_{\mathcal{C}}\boldsymbol{U}_{\mathcal{C}\rightarrow} = 0, \tag{5.22}$$

where the second equality is because $\boldsymbol{U}$ is row supported in $\mathcal{C}$. Thus, we have

$$\boldsymbol{G}_{\mathcal{C}}^{\dagger}\boldsymbol{G}\boldsymbol{R} = (\boldsymbol{G}_{\mathcal{C}}^{\star}\boldsymbol{G}_{\mathcal{C}})^{-1}\boldsymbol{G}_{\mathcal{C}}^{\star}\boldsymbol{G}\boldsymbol{R} = 0 \quad \text{and therefore} \quad \boldsymbol{G}_{\mathcal{C}}^{\star}\boldsymbol{G}\boldsymbol{R} = 0. \tag{5.23}$$

This shows that the residual $\boldsymbol{R}$ plays no role in the last term in (5.21),

$$\left\|\left(\boldsymbol{G}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon} - \boldsymbol{X} + \boldsymbol{R})\right)_{\mathcal{C}\rightarrow}\right\|_{1,2} = \left\|\left(\boldsymbol{G}^{\star}\boldsymbol{G}(\boldsymbol{X}^{\varepsilon} - \boldsymbol{X})\right)_{\mathcal{C}\rightarrow}\right\|_{1,2} = \left\|\left(\boldsymbol{G}^{\star}\boldsymbol{W}^{\varepsilon}\right)_{\mathcal{C}\rightarrow}\right\|_{1,2}, \tag{5.24}$$

as stated in Theorem 2.5. $\square$

The proof of Theorem 2.6 is exactly the same as in section 5.3. Note that the estimate (2.29) applies to the matrix $\boldsymbol{X}$ and its projection $\boldsymbol{U}$ and aggregation $\overline{\boldsymbol{X}}$ defined relative to the set $\mathcal{C}$, instead of $\mathcal{S}$. Thus, (2.50) follows directly from (2.29). $\square$

It remains to prove Lemma 2.4. Recall that $\boldsymbol{U}$ is defined via the projection (2.38). This induces a linear operator $\mathfrak{T}: \mathbb{C}^{N_r \times N_v} \to \mathbb{C}^{N_r \times N_v}$ that maps $\boldsymbol{G}\boldsymbol{X}$ to $\boldsymbol{G}\boldsymbol{U}$,

$$\boldsymbol{G}\boldsymbol{U} = \mathfrak{T}\boldsymbol{G}\boldsymbol{X}. \tag{5.25}$$

Note that $\boldsymbol{G}\boldsymbol{U} = \mathfrak{T}\boldsymbol{G}\boldsymbol{U}$ and since (5.23) gives

$$0 = \boldsymbol{G}_{\mathcal{C}}^{\star}\boldsymbol{G}\boldsymbol{R} = \boldsymbol{G}_{\mathcal{C}}^{\star}\boldsymbol{G}(\boldsymbol{X} - \boldsymbol{U}) = \boldsymbol{G}_{\mathcal{C}}^{\star}(\boldsymbol{G}\boldsymbol{X} - \mathfrak{T}\boldsymbol{G}\boldsymbol{X}), \tag{5.26}$$

$\mathfrak{T}$ is the orthogonal projection onto the range of $\boldsymbol{G}_{\mathcal{C}}$.

Our goal is to estimate

$$\|\boldsymbol{G}\boldsymbol{R}\|_F^2 = \|\boldsymbol{G}(\boldsymbol{X} - \boldsymbol{U})\|_F^2 = \sum_{v=1}^{N_v} \|\boldsymbol{G}(\boldsymbol{x}_v - \boldsymbol{u}_v)\|_2^2 = \sum_{v=1}^{N_v} \|\boldsymbol{G}\boldsymbol{x}_v - \mathfrak{T}\boldsymbol{G}\boldsymbol{x}_v\|_2^2, \tag{5.27}$$

where we note that since $\mathfrak{T}$ is the orthogonal projection on range($\boldsymbol{G}_{\mathcal{C}}$),

$$\|\boldsymbol{G}\boldsymbol{x}_v - \mathfrak{T}\boldsymbol{G}\boldsymbol{x}_v\|_2 \leq \|\boldsymbol{G}\boldsymbol{x}_v - \boldsymbol{z})\|_2, \quad \forall \boldsymbol{z} \in \text{range}(\boldsymbol{G}_{\mathcal{C}}). \tag{5.28}$$

We use this inequality for $\boldsymbol{z} = \boldsymbol{G}\overline{\boldsymbol{X}} = \boldsymbol{G}_{\mathcal{C}}\overline{\boldsymbol{X}}_{\mathcal{C}\rightarrow}$, and obtain

$$\|\boldsymbol{G}\boldsymbol{x}_v - \mathfrak{T}\boldsymbol{G}\boldsymbol{x}_v\|_2 \leq \|\boldsymbol{G}\boldsymbol{x}_v - \boldsymbol{G}\overline{\boldsymbol{x}_v})\|_2 = \left\|\sum_{j\in\mathcal{S}} X_{j,v}\boldsymbol{g}_j - \sum_{j\in\mathcal{C}} \overline{X_{j,v}}\boldsymbol{g}_j\right\|_2, \tag{5.29}$$

because $\boldsymbol{X}$ is row supported in $\mathcal{S}$ and $\overline{\boldsymbol{X}}$ is row supported in $\mathcal{C}$. Next, using the decomposition (2.35) of $\mathcal{S}$ and the definition (2.36) of $\overline{\boldsymbol{X}}$, we have

$$\left\|\sum_{j\in\mathcal{S}} X_{j,v}\boldsymbol{g}_j - \sum_{j\in\mathcal{C}} \overline{X_{j,v}}\boldsymbol{g}_j\right\|_2 = \left\|\sum_{j\in\mathcal{C}}\sum_{l\in\mathscr{S}_j} X_{l,v}\boldsymbol{g}_l - \sum_{j\in\mathcal{C}}\left[\sum_{l\in\mathscr{S}_j} X_{l,v}\mu(\boldsymbol{g}_j,\boldsymbol{g}_l)\right]\boldsymbol{g}_j\right\|_2$$

$$= \left\|\sum_{j\in\mathcal{C}}\sum_{l\in\mathscr{S}_j} X_{l,v}\left[\boldsymbol{g}_l - \mu(\boldsymbol{g}_j,\boldsymbol{g}_l)\boldsymbol{g}_j\right]\right\|_2. \tag{5.30}$$

We can bound this using the triangle inequality and

$$\left\|\boldsymbol{g}_l - \mu(\boldsymbol{g}_j,\boldsymbol{g}_l)\boldsymbol{g}_j\right\|_2^2 = \langle \boldsymbol{g}_l - \mu(\boldsymbol{g}_j,\boldsymbol{g}_l)\boldsymbol{g}_j, \boldsymbol{g}_l - \mu(\boldsymbol{g}_j,\boldsymbol{g}_l)\boldsymbol{g}_j\rangle = 1 - |\mu(j,l)|^2 \leq 2\mathfrak{D}(j,l), \tag{5.31}$$

where we used the definition of the semimetric $\mathfrak{D}$ and of μ. Since $\mathscr{S}_j$ is contained within a ball of radius r_c centered at $j \in \mathcal{C}$, we have $\mathfrak{D}(j,l) \leq r_c$ in (5.31), and gathering the results (5.29)–(5.31), we get

$$\|\boldsymbol{G}\boldsymbol{x}_v - \mathfrak{T}\boldsymbol{G}\boldsymbol{x}_v\|_2 \leq \sqrt{2r_c} \sum_{j \in \mathcal{C}} \sum_{l \in \mathscr{S}_j} |X_{l,v}| = \sqrt{2r_c}\|\boldsymbol{x}_v\|_1. \tag{5.32}$$

Finally, substituting in (5.27),

$$\|\boldsymbol{G}\boldsymbol{R}\|_F^2 \leq 2r_c \sum_{v=1}^{N_v} \left(\|\boldsymbol{x}_v\|_1\right)^2 = 2r_c\|\boldsymbol{X}^T\|_{2,1}^2, \tag{5.33}$$

as stated in the Lemma. $\square$

6. Summary. We presented a resolution theory for synthetic aperture radar (SAR) imaging using the multiple measurement vector (MMV) approach, also known as simultaneously sparse optimization. This seeks to find an unknown matrix $\boldsymbol{X}$ with sparse row support, by inverting a linear system of equations using sparsity promoting convex optimization. In the SAR imaging application, $\boldsymbol{X}$ models the unknown reflectivity of a scattering scene. The rows of $\boldsymbol{X}$ are indexed by the points in the imaging region, and the columns correspond to its values for multiple views of the imaging scene, from different sub-apertures and polarization diverse measurements.

The resolution theory does not pursue the question of exact recovery, but seeks to estimate the neighborhood of the support of $\boldsymbol{X}$ where the largest entries in the reconstruction lie. The radius of this neighborhood represents the resolution limit and it depends on the noise level. We introduced a quantifier of how the unknowns influence each other in imaging, called the multiple view interaction coefficient, and showed that the smaller this is and the weaker the noise, the better the estimate of the support of $\boldsymbol{X}$. We also quantified the error of the reconstruction and studied the advantage of having multiple views. The existing literature shows that the MMV method does not always perform better than sparsity promoting optimization with a single view, the so-called single measurement vector (SMV) formulation. We showed that if the rows of $\boldsymbol{X}$ are orthogonal, then the MMV approach is expected to perform better, depending on how the unknowns are distributed in the imaging scene. We quantified this advantage and explained how the condition of orthogonality of the rows of $\boldsymbol{X}$ arises in the application of SAR imaging of direction dependent reflectivity.

We also studied imaging of well-separated clusters of scatterers and showed that the MMV approach gives a reconstruction supported near these clusters, with values near the net reflectivity of the clusters.

Appendix A. Proof of Proposition 3.1. We begin with the Taylor expansion of the exponent in (3.21) with respect to the arclength r

$$\overline{k}\frac{(\boldsymbol{r}_o + r\boldsymbol{\tau}_1 - \overline{\boldsymbol{y}})}{|\boldsymbol{r}_o + r\boldsymbol{\tau}_1 - \overline{\boldsymbol{y}}|} \cdot (\boldsymbol{y}_q - \boldsymbol{y}_l) = \overline{k}\boldsymbol{m}_o \cdot (\boldsymbol{y}_q - \boldsymbol{y}_l) + \overline{k}r\frac{\boldsymbol{\tau}_1 \cdot \mathbb{P}_o(\boldsymbol{y}_q - \boldsymbol{y}_l)}{|\boldsymbol{r}_o - \overline{\boldsymbol{y}}|} + \ldots, \tag{A.1}$$

with $\boldsymbol{m}_o$ and $\mathbb{P}_o$ defined in (3.24). We suppose that the aperture A and the cross-range offset between $\boldsymbol{y}_q$ and $\boldsymbol{y}_l$ are small enough so we can neglect the higher terms[‡‡] in (A.1). Then, (3.21) may be estimated by

$$|\mu(\boldsymbol{x}_{q\rightarrow}, \boldsymbol{x}_{l\rightarrow})| \approx \frac{A\left|\psi_{q,l}(A/2)e^{iQ/2} - \psi_{q,l}(-A/2)e^{-iQ/2} - \int_{-A/2}^{A/2} dr\,\psi_{q,l}'(r)e^{irQ/A}\right|}{|Q|\|\psi_{q,q}\|_{L_1(-A/2,A/2)}^{1/2}\|\psi_{l,l}\|_{L_1(-A/2,A/2)}^{1/2}}. \tag{A.2}$$

If the reflectivities are independent of direction, (A.2) becomes $|\mu(\boldsymbol{x}_{q\rightarrow}, \boldsymbol{x}_{l\rightarrow})| \approx |\mathrm{sinc}(Q/2)|$. This attains its maximum at $Q = 0$ i.e., at $q = l$, and decays as $1/|Q|$, as stated in the proposition. It remains to show

---

[‡‡]The results are qualitatively the same if we include quadratic terms in r and neglect cubic and higher order terms. The discussion is simpler if we consider only the shown terms in (A.1).

that the result extends to reflectivities that vary smoothly with direction. We obtain from (A.2), using the triangle inequality, that

$$|\mu(\boldsymbol{x}_{q\to}, \boldsymbol{x}_{l\to})| \leq \frac{A\big[|\psi_{q,l}(A/2)| + |\psi_{q,l}(-A/2)| + \|\psi'_{q,l}\|_{L_1(-A/2,A/2)}\big]}{|Q|\|\|\psi_{q,q}\|_{L_1(-A/2,A/2)}^{1/2}\|\|\psi_{l,l}\|_{L_1(-A/2,A/2)}^{1/2}}, \tag{A.3}$$

and we estimate next the three terms in the numerator.

We begin with $|\psi_{q,l}(A/2)$, which satisfies

$$|\psi_{q,l}(A/2)| \leq |\psi_{q,l}(s)| + \Big| \int_s^{A/2} dr\, \psi'_{q,l}(r) \Big|, \tag{A.4}$$

by the fundamental theorem of calculus and the triangle inequality. Therefore,

$$\begin{aligned}
A|\psi_{q,l}(A/2)| &= \int_{-A/2}^{A/2} ds\, |\psi_{q,l}(A/2)| \\
&\leq \int_{-A/2}^{A/2} ds\, |\psi_{q,l}(s)| + \int_{-A/2}^{A/2} ds \Big| \int_s^{A/2} dr\, \psi'_{q,l}(r) \Big| \\
&\leq \|\psi_{q,l}\|_{L_1(-A/2,A/2)} + \int_{-A/2}^{A/2} ds \int_{-A/2}^{A/2} dr\, |\psi'_{q,l}(r)| \\
&= \|\psi_{q,l}\|_{L_1(-A/2,A/2)} + A\|\psi'_{q,l}\|_{L_1(-A/2,A/2)}.
\end{aligned} \tag{A.5}$$

The first term in this equation can be bound using the Cauchy-Schwartz inequality, once we recall the definition (3.22) of $\psi_{q,l}$. We rewrite this definition as

$$\psi_{q,l}(r) = \xi_q(\boldsymbol{r}_o + r\boldsymbol{\tau}_1)\xi_l^\star(\boldsymbol{r}_o + r\boldsymbol{\tau}_1), \tag{A.6}$$

in an abuse of notation, so that

$$\xi_{q,v} = \xi_q(\boldsymbol{r}_v), \quad \boldsymbol{r}_v = \boldsymbol{r}_o + \Big(\frac{v-1}{N_v-1} - \frac{1}{2}\Big)A\boldsymbol{\tau}_1.$$

Then, we have

$$\begin{aligned}
L_2\|\psi_{q,l}\|_{L_1(-A/2,A/2)} &= \int_{-A/2}^{A/2} dr\, |\xi_q(\boldsymbol{r}_o + r\boldsymbol{\tau}_1)\xi_l^\star(\boldsymbol{r}_o + r\boldsymbol{\tau}_1)| \\
&\leq \Big[\int_{-A/2}^{A/2} dr\, |\xi_q(\boldsymbol{r}_o + r\boldsymbol{\tau}_1)|^2\Big]^{1/2} \Big[\int_{-A/2}^{A/2} dr\, |\xi_l(\boldsymbol{r}_o + r\boldsymbol{\tau}_1)|^2\Big]^{1/2} \\
&= \|\psi_{q,q}\|_{L_1(-A/2,A/2)}^{1/2}\|\psi_{l,l}\|_{L_1(-A/2,A/2)}^{1/2}.
\end{aligned} \tag{A.7}$$

We also have from (A.6) that

$$\psi'_{q,l}(r) = \boldsymbol{\tau}_1 \cdot \nabla\xi_q(\boldsymbol{r}_o + r\boldsymbol{\tau}_1)\xi_l^\star(\boldsymbol{r}_o + r\boldsymbol{\tau}_1) + \xi_q(\boldsymbol{r}_o + r\boldsymbol{\tau}_1)\boldsymbol{\tau}_1 \cdot \nabla\xi_l^\star(\boldsymbol{r}_o + r\boldsymbol{\tau}_1), \tag{A.8}$$

and obtain using the Cauchy-Schwartz and triangle inequalities that

$$\begin{aligned}
\|\psi'_{q,l}\|_{L_1(-A/2,A/2)} &\leq \|\boldsymbol{\tau}_1 \cdot \nabla\xi_q\|_{L_2(-A/2,A/2)}\|\xi_l\|_{L_2(-A/2,A/2)} + \|\xi_q\|_{L_2(-A/2,A/2)}\|\boldsymbol{\tau}_1 \cdot \nabla\xi_l\|_{L_2(-A/2,A/2)} \\
&\leq \|\nabla\xi_q\|_{L_2(-A/2,A/2)}\|\xi_l\|_{L_2(-A/2,A/2)} + \|\xi_q\|_{L_2(-A/2,A/2)}\|\nabla\xi_l\|_{L_2(-A/2,A/2)}.
\end{aligned} \tag{A.9}$$

To estimate this further, let us introduce the constant K_q, which depends on the scale of variation of the reflectivity ξ_q, such that

$$\|\nabla\xi_q\|_{L_2(-A/2,A/2)} \leq \frac{K_q}{A}\|\xi_q\|_{L_2(-A/2,A/2)}. \tag{A.10}$$

Since $\|\xi_q\|_{L_2(-A/2,A/2)} = \|\psi_{q,q}\|_{L_1(-A/2,A/2)}^{1/2}$, by definition (A.6), we obtain from (A.9)–(A.10) that

$$A\|\psi'_{q,l}\|_{L_1(-A/2,A/2)} \leq (K_q + K_l)\|\psi_{q,q}\|_{L_1(-A/2,A/2)}^{1/2}\|\psi_{l,l}\|_{L_1(-A/2,A/2)}^{1/2}. \tag{A.11}$$

The estimate (A.5) becomes

$$A|\psi_{q,l}(A/2)| \leq (1 + K_q + K_l)\|\psi_{q,q}\|_{L_1(-A/2,A/2)}^{1/2}\|\psi_{l,l}\|_{L_1(-A/2,A/2)}^{1/2}, \tag{A.12}$$

and a similar bound applies to $A|\psi_{q,l}(-A/2)|$.

Gathering the results (A.11)–(A.12) and substituting in (A.3), we obtain the statement of the proposition, with $C_{q,l} = 12\pi(1 + K_q + K_q)$. $\square$

Appendix B. Expression of matrix $\mathbf{\Gamma}(\overline{r})$. The 6×6 matrix $\mathbf{\Gamma}(\overline{r})$ that enters the data model (4.9) can be written as

$$\mathbf{\Gamma}(\overline{r}) = \mathbf{\Gamma}_{\text{diag}}(\overline{r}) + \mathbf{\Gamma}_{\text{off-diag}}(\overline{r})$$

where

$$\mathbf{\Gamma}_{\text{diag}}(\overline{r}) = \text{diagonal}\Big((1 - \eta_1^2)^2, (1 - \eta_2)^2, (1 - \beta^2)^2, (1 - \eta_1^2)(1 - \eta_2^2) + (\eta_1\eta_2)^2,$$

$$(1 - \eta_1^2)(1 - \beta^2) + (\eta_1\beta)^2, (1 - \eta_2^2)(1 - \beta^2) + (\eta_2\beta)^2\Big),$$

is the diagonal part of $\mathbf{\Gamma}(\overline{r})$ and

$$\mathbf{\Gamma}_{\text{off-diag}}(\overline{r}) = \begin{pmatrix} 0 & (\eta_1\eta_2)^2 & (\eta_1\beta)^2 & \eta_1\eta_2(\eta_2^2 - 1) & \eta_1\beta(\eta_1^2 - 1) & \eta_1^2\eta_2\beta \\ (\eta_1\eta_2)^2 & 0 & (\eta_2\beta)^2 & \eta_1\eta_2(\eta_2^2 - 1) & \eta_1\eta_2^2\beta & \eta_2\beta(\eta_2^2 - 1) \\ (\eta_1\beta)^2 & (\eta_2\beta)^2 & 0 & \eta_1\eta_2\beta^2 & \eta_1\beta(\beta^2 - 1) & \eta_2\beta(\beta^2 - 1) \\ 2\eta_1\eta_2(\eta_1^2 - 1) & 2\eta_1\eta_2(\eta_2^2 - 1) & 2\eta_1\eta_2\beta^2 & 0 & \eta_2\beta(2\eta_1^2 - 1) & \eta_1\beta(2\eta_2^2 - 1) \\ 2\eta_1\beta(\eta_1^2 - 1) & 2\eta_1\eta_2^2\beta & 2\eta_1\beta(\beta^2 - 1) & \eta_2\beta(2\eta_1^2 - 1) & 0 & \eta_1\eta_2(2\beta^2 - 1) \\ 2\eta_1^2\eta_2\beta & 2\eta_2\beta(\eta_2^2 - 1) & 2\eta_2\beta(\beta^2 - 1) & \eta_1\beta(2\eta_2^2 - 1) & \eta_1\eta_2(2\beta^2 - 1) & 0 \end{pmatrix}$$

is its off-diagonal part.

REFERENCES

[1] *GOTCHA volumetric SAR data set.* https://www.sdms.afrl.af.mil/index.php?collection=gotcha.
[2] Habib Ammari, Josselin Garnier, Wenjia Jing, Hyeonbae Kang, Mikyoung Lim, Knut Sølna, and Han Wang, *Mathematical and statistical methods for multistatic imaging*, vol. 2098, Springer, 2013.
[3] Habib Ammari, Ekaterina Iakovleva, Dominique Lesselier, and Gaële Perrusson, *MUSIC-type electromagnetic imaging of a collection of small three-dimensional inclusions*, SIAM Journal on Scientific Computing, 29 (2007), pp. 674–709.
[4] Richard Baraniuk and Philippe Steeghs, *Compressive radar imaging*, in Radar Conference, 2007 IEEE, IEEE, 2007, pp. 128–133.
[5] B. Biondi, *3D seismic imaging*, Society of Exploration Geophysicists, 2006.
[6] Jérôme Bobin, Jean-Luc Starck, and Roland Ottensamer, *Compressed sensing in astronomy*, IEEE Journal of Selected Topics in Signal Processing, 2 (2008), pp. 718–726.
[7] Liliana Borcea and Ilker Kocyigit, *Resolution analysis of imaging with ℓ_1 optimization*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 3015–3050.
[8] Liliana Borcea, Miguel Moscoso, George Papanicolaou, and Chrysoula Tsogka, *Synthetic aperture imaging of direction-and frequency-dependent reflectivities*, SIAM Journal on Imaging Sciences, 9 (2016), pp. 52–81.
[9] A. M. Bruckstein, D. L. Donoho, and M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM review, 51 (2009), pp. 34–81.
[10] Emmanuel J Candès, Justin Romberg, and Terence Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on information theory, 52 (2006), pp. 489–509.
[11] E. J. Candes and T. Tao, *Decoding by linear programming*, Information Theory, IEEE Transactions on, 51 (2005), pp. 4203–4215.
[12] Emmanuel J Candes and Terence Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies?*, IEEE transactions on information theory, 52 (2006), pp. 5406–5425.
[13] A. Chai, M. Moscoso, and G. Papanicolaou, *Robust imaging of localized scatterers using the singular value decomposition and 1 minimization*, Inverse Problems, 29 (2013), p. 025016.

[14] ———, *Imaging strong localized scatterers with sparsity promoting optimization*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 1358–1387.

[15] J. Chen and X. Huo, *Theoretical results on sparse representations of multiple-measurement vectors*, IEEE Transactions on Signal Processing, 54 (2006), pp. 4634–4643.

[16] Margaret Cheney, *Imaging frequency-dependent reflectivity from synthetic-aperture radar*, Inverse Problems, 29 (2013), p. 054002.

[17] Margaret Cheney and Brett Borden, *Fundamentals of radar imaging*, SIAM, 2009.

[18] Albert Cohen, Wolfgang Dahmen, and Ronald Devore, *Compressed sensing and best k-term approximation*, J. Amer. Math. Soc, (2009), pp. 211–231.

[19] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, *Sparse solutions to linear inverse problems with multiple measurement vectors*, Signal Processing, IEEE Transactions on, 53 (2005), pp. 2477–2488.

[20] John C Curlander and Robert N McDonough, *Synthetic aperture radar*, John Wiley & Sons New York, NY, USA, 1991.

[21] CVX Research, *Cvx: matlab software for disciplined convex programming, version 2.0.* http://cvxr.com/cvxhttp://cvxr.com/cvx, August 2012.

[22] David L Donoho, *Compressed sensing*, IEEE Transactions on information theory, 52 (2006), pp. 1289–1306.

[23] David L Donoho and Michael Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 2197–2202.

[24] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inf. Theor., 47 (2006), pp. 2845–2862.

[25] David L Donoho and Benjamin F Logan, *Signal recovery and the large sieve*, SIAM Journal on Applied Mathematics, 52 (1992), pp. 577–591.

[26] David L Donoho and Philip B Stark, *Uncertainty principles and signal recovery*, SIAM Journal on Applied Mathematics, 49 (1989), pp. 906–931.

[27] Y. C. Eldar and M. Mishali, *Robust recovery of signals from a structured union of subspaces*, IEEE Transactions on Information Theory, 55 (2009), pp. 5302–5316.

[28] A. Fannjiang and H-C Tseng, *Compressive radar with off-grid targets: a perturbation approach*, Inverse Problems, 29 (2013), p. 054008.

[29] A. C. Fannjiang, T. Strohmer, and P. Yan, *Compressed remote sensing of sparse objects*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 595–618.

[30] Michael Lustig, David Donoho, and John M Pauly, *Sparse mri: The application of compressed sensing for rapid mr imaging*, Magnetic resonance in medicine, 58 (2007), pp. 1182–1195.

[31] D. Malioutov, M. Cetin, and A. S. Willsky, *A sparse signal reconstruction perspective for source localization with sensor arrays*, IEEE Transactions on Signal Processing, 53 (2005), pp. 3010–3022.

[32] Lee C Potter, Emre Ertin, Jason T Parker, and Müjdat Cetin, *Sparsity and compressed sensing in radar imaging*, Proceedings of the IEEE, 98 (2010), pp. 1006–1020.

[33] Fadil Santosa and William W Symes, *Linear inversion of band-limited reflection seismograms*, SIAM Journal on Scientific and Statistical Computing, 7 (1986), pp. 1307–1330.

[34] Paul Sotirelis, Jason Parker, Xueyu Hu, Margaret Cheney, and Matthew Ferrara, *Frequency-dependent reflectivity image reconstruction*, in SPIE Defense, Security, and Sensing, International Society for Optics and Photonics, 2013, pp. 874602–874602.

[35] Paul Sotirelis, Jason T Parker, Michael Fu, Xueyu Hu, and Richard Albanese, *A study of material identification using sar*, in Radar Conference (RADAR), 2012 IEEE, IEEE, 2012, pp. 0112–0115.

[36] J. A. Tropp, *Greed is good: Algorithmic results for sparse approximation*, Information Theory, IEEE Transactions on, 50 (2004), pp. 2231–2242.

[37] Joel A. Tropp, *Algorithms for simultaneous sparse approximation: Part ii: Convex relaxation*, Signal Process., 86 (2006), pp. 589–602.

[38] Joel A. Tropp, Anna C. Gilbert, and Martin J. Strauss, *Algorithms for simultaneous sparse approximation: Part i: Greedy pursuit*, Signal Process., 86 (2006), pp. 572–588.

[39] E. van den Berg and M. P. Friedlander, *Theoretical and empirical results for recovery from multiple measurements*, IEEE Transactions on Information Theory, 56 (2010), pp. 2516–2527.

[40] Yuehao Wu, Peng Ye, Iftekhar O Mirza, Gonzalo R Arce, and Dennis W Prather, *Experimental demonstration of an optical-sectioning compressive sensing microscope (csm)*, Optics express, 18 (2010), pp. 24565–24578.