

Synthetic aperture imaging and motion estimation using tensor methods

Matan Leibovich*, George Papanicolaou† and Chrysoula Tsogka‡

Abstract

We consider a synthetic aperture imaging configuration, such as synthetic aperture radar (SAR), where we want to first separate reflections from moving targets from those coming from a stationary background, and then to image separately the moving and the stationary reflectors. For this purpose, we introduce a representation of the data as a third order tensor formed from data coming from partially overlapping subapertures. We then apply a tensor robust principal component analysis (TRPCA) to the tensor data which separates them into the parts coming from the stationary and moving reflectors. A key feature of the proposed algorithm is the use of a Fourier based tensor nuclear norm which is well adapted to the SAR data structure. Images are then formed with the separated data sets. Our analysis shows a distinctly improved performance of TRPCA, compared to the usual matrix case. In particular, the tensor decomposition can identify motion features that are undetectable when using the conventional motion estimation methods, including matrix RPCA. We illustrate the performance of the method with numerical simulations in the X-band radar regime.

1 Introduction

The problem of separating the echoes of moving targets from those of a stationary background in synthetic aperture radar (SAR) imaging is important because different imaging methods need to be employed in each case. We show that the separation is improved considerably when the data is recast in tensor form, and the convex optimization problem of tensor robust principal component analysis (TRPCA) is used. TRPCA requires the generalization of matrix norms to tensor form, the challenge being to use a suitable tensor nuclear norm. We use a Fourier based tensor nuclear norm which can capture the motion over multiple scales of variation and therefore it is well suited for the SAR data structure. We first review briefly SAR imaging, then the application of RPCA for motion detection, and follow with a statement of the main results of this paper.

*Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305. (matanle@stanford.edu)

†Department of Mathematics, Stanford University, Stanford, CA 94305. (papanicolou@stanford.edu)

‡Department of Applied Mathematics, University of California, Merced, 5200 North Lake Road, Merced, CA 95343 (ct-sogka@ucmerced.edu)

1.1 The synthetic aperture radar imaging problem

Synthetic aperture radar is used extensively in satellite and airborne imaging for many different applications [29]. The main idea behind SAR is to combine coherently the information obtained with a single transmitter-receiver that is probing the medium from multiple locations. Thus, one can form a synthetic aperture and achieve high resolution images of reflectivity, even though the single receiver is incapable of resolving the scattered wavefronts. This is particularly relevant for airborne radar platforms.

Data is collected by a moving platform, with a *slow-time* s dependent position $\vec{r}(s)$, emitting a sequence of *fast-time* broadband pulses $f(t)$, and recording the echoes corresponding to each pulse. The pulses have a limited time support, with a pulse repetition interval Δs , so that echoes from different pulses do not overlap. The data collected are denoted by $D(s, t)$, the t -dependent series of echoes received from a pulse transmitted at $\vec{r}(s)$. A schematic of a SAR imaging configuration is shown in Figure 1.

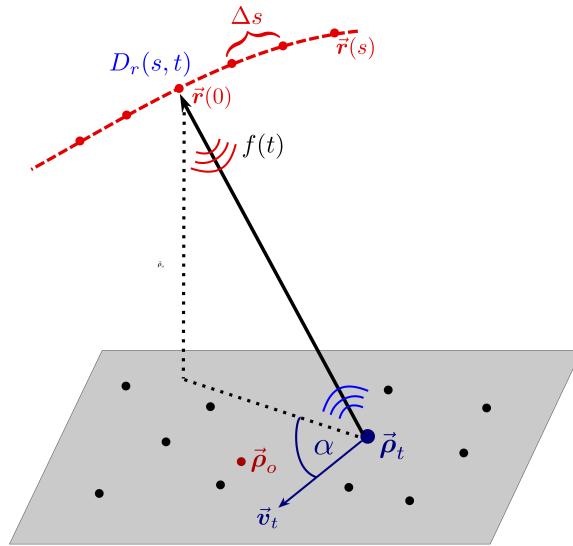


Figure 1: Synthetic aperture imaging configuration. At every time s the airborne platform emits a pulse $f(t)$ and records the reflections $D(t, s)$ coming from the imaging region. There are 10 stationary point scatterers, and a single moving target. The angle α in which the target is moving is relative to the vertical platform-target plane at $s = 0$, $\vec{r}(0)$. Here \vec{p}_o is a fixed reference point in the imaging region, used to compress the support of the data (1.2)

To maximize the power emitted, the probing pulses are long, of support $t_c \gg 1/B$ where B denotes the bandwidth. They are linear frequency modulated chirps. To re-concentrate the energy of the reflected echoes to an interval of size $1/B$ they are convolved with the complex conjugate of the time-reversed emitted pulse. This is the pulse compression step. Since the reflections relevant for imaging cover a limited area of support much smaller than Δs , a range compression is done as well, that is, we remove from the data the large phase $\omega\tau(s, \vec{p}_o)$ where \vec{p}_o is a reference point, constant during the data acquisition. Pulse and range compression

together give the down-ramped data,

$$D_r(s, t) = \int dt' D(s, t - t' + \tau(s, \vec{\rho}_o)) \overline{f(-t')} = \int \frac{d\omega}{2\pi} \overline{\widehat{f}(\omega)} \widehat{D}(s, \omega) e^{-i\omega[t + \tau(s, \vec{\rho}_o)]}. \quad (1.1)$$

Here, $\tau(s, \vec{\rho}_o)$ is the round-trip travel time between the platform location at slow time s , $\vec{r}(s)$ and the reference point location $\vec{\rho}_o$,

$$\tau(s, \vec{\rho}_o) = 2 \frac{\|\vec{r}(s) - \vec{\rho}_o\|}{c}, \quad (1.2)$$

with c the speed of light.

The SAR data matrix $\mathcal{D} \in \mathbb{R}^{(n+1) \times (m+1)}$ is actually obtained in discrete samples of $D_r(s, t)$.

$$\mathcal{D}_{il} = D_r(s_{i-\frac{n}{2}-1}, t_{l-1}), \quad i = 1, \dots, n+1, \quad l = 1, \dots, m+1, \quad (1.3)$$

with slow times s_j defined by

$$s_j = j\Delta s, \quad j = -n/2, \dots, n/2 \quad (1.4)$$

and fast times t_l defined as

$$t_l = l\Delta t, \quad l = 1, \dots, m. \quad (1.5)$$

Here we assumed that the pulse repetition rate Δs is an integer multiple of Δt and set $m = \Delta s / \Delta t$. The SAR image is formed by summing coherently the down-ramped data $D_r(s_j, t)$ back-propagated to the imaging point $\vec{\rho}$ using the travel times differences $\tau(s_j, \vec{\rho}) - \tau(s_j, \vec{\rho}_o)$,

$$I^{SAR}(\vec{\rho}) = \sum_{j=-n/2}^{n/2} D_r(s_j, \tau(s_j, \vec{\rho}) - \tau(s_j, \vec{\rho}_o)). \quad (1.6)$$

The SAR image processing (1.6) assumes that only reflections from stationary targets are contained in the down-ramped data $D_r(s, t)$. Consequently, if moving targets, as illustrated in Figure 1, are present in the region to be imaged their reflections are not correctly back-propagated and this results in blurred images affected by the reflectivity and the velocity of the moving targets. For a complex scene with many stationary and moving targets the image may be severely distorted and neither the stationary nor the moving targets may be imaged or tracked. To address this issue several motion estimation and separation strategies have been developed, which we now review briefly.

The oldest and most widely used approach is the Displaced Phase Center Antenna (DPCA) method in which two synchronized antennas are used, following the same trajectory with a small time delay. By subtracting the data traces collected at the two antennas the echoes due to the stationary background are essentially eliminated. This approach does not aim at image formation as part of motion detection, but requires the necessary hardware to be in place to record the extra data. We refer to the classical handbook on SAR [34] and a review report from the Lincoln Laboratory on this technique [30], which has been used to improve the performance of moving-target-indicators radars since the 1950s.

Other well known approaches are autofocus based algorithms [12, 2, 23] or more generally space-time adaptive processing algorithms [11, 36]. For a recent review on sparsity driven techniques for SAR imaging of scenes containing moving objects we refer to [8]. All these algorithms for moving target detection rely on

forming a preliminary image first, and then detecting motion by sharpening features in the image.

In this paper we take a different approach and rely on robust principal component analysis (RPCA) to solve the SAR data separation problem. RPCA has been extensively used in SAR in various applications. In [38] the authors perform the principal component analysis in the range doppler domain. As they point out the choice of the weighting parameter η (see (1.7)) is important and affects the performance of the algorithm. Other works, such as [31, 39], utilize low-rank and sparsity assumptions in the formed image. In this case images corresponding to different subapertures in the azimuth direction are first computed. These images are stacked as columns and form a matrix on which low rank plus sparse decomposition is performed. This is the classical way of using low rank plus sparse decomposition as in the video applications. In [33, 27], sparsity in the formed image is imposed using priors and a sparse dictionary representation.

Our approach exploits properties of the raw-data matrix so as to detect motion. It does not require the formation of the SAR image or the use of any special hardware. Only the standard monostatic single transmitter/receiver SAR data are used. Robust principal component analysis is traditionally used to separate signals from noise. Here we use it to decompose a matrix into its low rank and sparse parts. The sparse part is not noise as in traditional RPCA [7, 41] but it is the signal corresponding to the moving targets echoes. Let us also refer to [3] where the authors develop a space time adaptive processing method that uses a projection onto the low rank part of a tensor data structure to reduce clutter in a multiple input multiple output (MIMO) radar system. Their approach uses a multi linear SVD (MLSVD) denoising method. In this work only the data from the moving target are considered as signal, the rest is noise. The idea of using RPCA for SAR data separation was first proposed in [38], [5] and further developed and analyzed in [22] where optimal parameters were derived for achieving robust separation in SAR. We explain next how RPCA can be used in SAR.

1.2 Robust principal component analysis for SAR data separation

RPCA, or low rank plus sparse decomposition, was originally applied in video processing [7]. RPCA uses the fact that the moving targets and stationary background would generate data structures with different spectral properties. Indeed, the background data form a low rank matrix, while the moving objects echoes correspond to a sparse matrix. These *data structures* can be decomposed by solving the convex optimization problem

$$\begin{aligned} \min_{L, S \in \mathbb{C}^{n_1 \times n_2}} \quad & \|L\|_* + \eta \|S\|_1 \\ \text{subject to} \quad & L + S = D. \end{aligned} \tag{1.7}$$

Here $\|L\|_*$ denotes the nuclear norm, that is the sum of the singular values of L , and $\|S\|_1$ is the matrix ℓ_1 -norm of S . Assuming the matrix $D \in \mathbb{R}^{n_1 \times n_2}$ is the sum of a low rank matrix, L_o , and a sparse matrix, S_o , then, under some additional conditions (cf. [7]), this optimization (1.7) recovers L_o and S_o exactly.

RPCA has found applications in a variety of problems in imaging and image processing, such as denoising, feature extraction, and data recovery [15, 28, 37]. The main idea in the application of RPCA to the SAR problem is that one can identify the stationary background as the low rank component of the SAR data matrix, and the moving targets as the sparse component [4].

In [22], we explored the performance of RPCA for the SAR data problem. We showed that one can

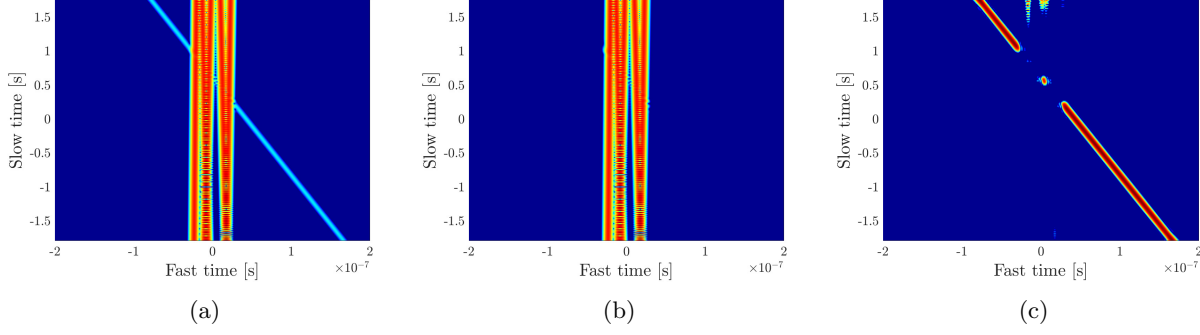


Figure 2: Example of RPCA applied to the data matrix $D_r(s, t)$. The moving target has velocity $\mathbf{v}_t = 15\text{m/s}$, $\alpha = 0$ and there are 10 stationary targets. The moving target's reflectivity is 10% of the reflectivity of the stationary scatterers. (a) Original data matrix. The echoes corresponding to the moving target is weak but still visible in the data; (b) RPCA: the low rank part for the stationary background; (c) RPCA: the sparse part for the echoes from the moving target. We observe that good separation is achieved with RPCA when using the optimal value for η in (1.7). This is the case for fast moving targets as in this example.

determine optimal parameters for achieving robust separation. An example of SAR data separation achieved using RPCA is illustrated in Figure 2. Here we use the simulation setup shown in Figure 1 with the moving target velocity $\mathbf{v}_t = 15\text{m/s}$ and $\alpha = 0$. The moving target's reflectivity is 10% of the reflectivity of the stationary scatterers. We observe that good separation is achieved with RPCA when using the optimal value for η in (1.7). This is the case for fast moving targets as in this example. RPCA has proven to be an efficient way to detect and separate moving targets in SAR data. The nuclear norm is a good indicator of motion, since the data traces associated with moving targets are supported over a larger number of columns, compared to stationary ones. However, the algorithm has its limitations, which we discuss next.

The column support of the target's echoes is determined by the possible values that the travel time difference, $\Delta\tau(s_j) = \tau(s_j, \vec{\rho}) - \tau(s_j, \vec{\rho}_o)$, takes in the aperture. In some cases, motion does not necessarily translate to an increased column support. More precisely, it was shown in [22] that the performance of RPCA depends on $N(\vec{\mathbf{v}}_t)$, which is an estimate of the number of columns spanned by the target's echoes. To first order, we can approximate $N(\vec{\mathbf{v}}_t)$ by

$$\begin{aligned} N(\vec{\mathbf{v}}_t) &\approx \frac{4S}{\Delta t} \frac{\vec{\mathbf{r}}(0) - \vec{\rho}_o}{\|\vec{\mathbf{r}}(0) - \vec{\rho}_o\|} \cdot \frac{\vec{\mathbf{v}}_t}{c} \\ &= \frac{4S}{\Delta t} \frac{\mathbf{v}_t}{c} \cos \alpha, \end{aligned} \quad (1.8)$$

where S is the total slow time aperture size. We observe that $N(\vec{\mathbf{v}}_t)$ depends not only on the magnitude of the moving target's velocity, \mathbf{v}_t , but also on the relative direction in which the target is moving, with respect to the platform, α . The direction $\alpha = 0$ is for targets moving parallel to the direction to the platform. Their traces will exhibit the largest variance in the value of $\Delta\tau(s)$. The direction $\alpha = \pi/2$ describes targets moving perpendicular to the direction to the platform, and their $\Delta\tau(s)$ will show much less variance as a function of s .

As can be seen in Figure 3, targets moving at directions with $\alpha \neq 0$ that exhibit smaller variations in the values $\Delta\tau(s)$ will be harder to detect using matrix RPCA. On the other hand, the associated data traces

still behave differently than the ones of a stationary background. Specifically we can see that the traces are non linear.

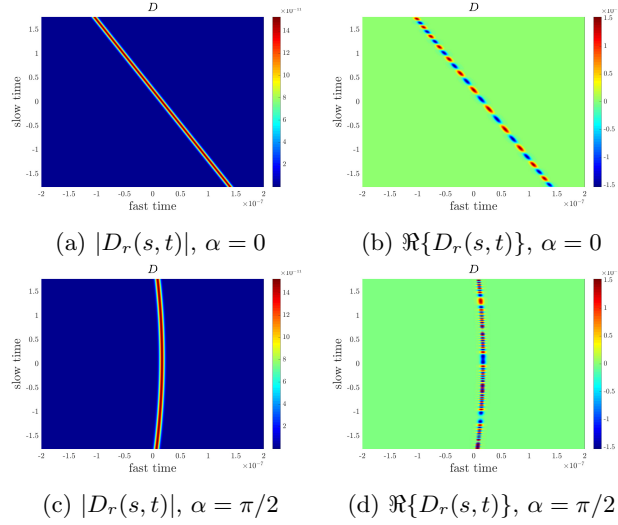


Figure 3: Effect of the direction of motion for a target moving at $\mathbf{v}_t = 15m/s$. When $\alpha = 0$, top, the column support is large, and the phase ($\Delta\tau(s)$) is nearly linear. With $\alpha = \pi/2$, bottom, the column support is much smaller, but the phase is non linear.

1.3 Main result of the paper: TRPCA for SAR data

The limitations of RPCA for motion estimation in SAR motivates a new look into possible extensions and modifications of the algorithm. Tensor based methods have been of great interest because high dimensional data arise naturally as tensors in more and more applications. Specifically in signal processing, tensor based methods for denoising and feature extraction are being used extensively [9]. For the SAR problem, this is motivated by the possibility of detecting and estimating more complex moving target behavior, such as non-linearity in the phase, by representing the data in a higher dimension. This is achieved here by dividing the large synthetic aperture into smaller, overlapping subapertures.

The problem of imaging moving targets in SAR can also be viewed as an image registration process. That is, a process that provides a precise correspondence between two or more images of the same object captured from different locations, at different times, or using different sensors. From this perspective, the introduction of subapertures for data processing is a natural one.

In this paper we use the SAR data, and the choice of subaperture and overlap size will be motivated differently. We recast the SAR data matrix as a third order tensor \mathcal{A} by dividing the large synthetic aperture into smaller overlapping subapertures, indexed by ℓ

$$\begin{aligned}
 A^{(\ell)}(s, t) &= D_r(s + \ell\vartheta s_{\text{sub}}, t), \quad s \in [0, s_{\text{sub}}], \\
 A^{(\ell)} &= \mathcal{A}(\cdot, \cdot, \ell) \in \mathbb{R}^{n_1 \times n_2}, \quad \ell = 0, \dots, n_3 - 1.
 \end{aligned} \tag{1.9}$$

The subapertures are with respect to the slow time s so that each subaperture contains the echoes from a subgroup of pulses. Here s_{sub} denotes the subaperture size, and ϑ is the overlap size, a number between

0 and 1. An illustration of tensor representation of the SAR data is given in Figure 4. The value of the hyper-parameters that define the tensor representation, i.e. s_{sub} and ϑ , affect the performance of TRPCA. A detailed analysis that allows us to determine the optimal value for s_{sub} and ϑ is carried out in Section 3. We then define the TRPCA algorithm, using a specific extension of the nuclear norm for third-order tensors

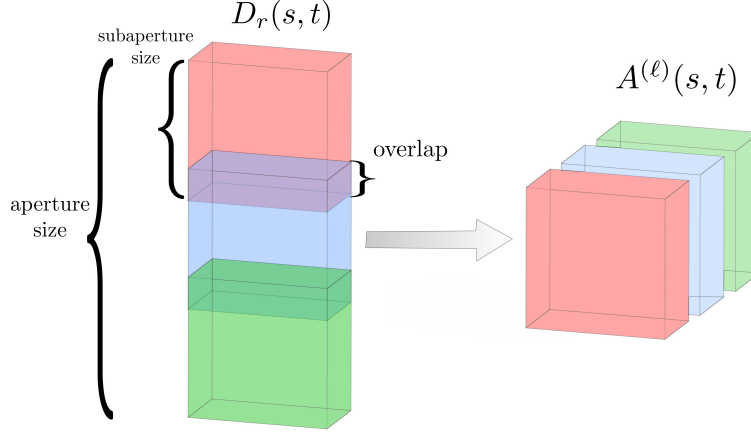


Figure 4: Schematic of SAR data tensor representation. For motion detection purposes, the large aperture data matrix $D_r(s, t)$ is converted to a 3rd order tensor $A^{(\ell)}(s, t)$. The tensor is composed of partially overlapping subapertures of the data $D_r(s, t)$.

[26], and solve a tensor based RPCA optimization problem for complex valued third order tensors:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S} \in \mathbb{C}^{n_1 \times n_2 \times n_3}} \quad & \|\mathcal{L}\|_{*, \mathcal{F}} + \eta \|\mathcal{S}\|_1, \\ \text{subject to} \quad & \mathcal{L} + \mathcal{S} = \mathcal{A} \end{aligned} \quad (1.10)$$

$\|\cdot\|_{*, \mathcal{F}}$, defined in (2.20), is a specific extension of the tensor nuclear norm which involves performing a Fourier Transform with respect to the subaperture index ℓ .

To evaluate the performance of the algorithm, we compare the separation achieved by matrix and tensor based RPCA. While the performance of TRPCA is not universally better than that of regular matrix RPCA, TRPCA performs significantly better in the cases where motion is hardest to detect as illustrated in the example considered in Figure 5. Here the setup shown in Figure 1 is considered with a slowly moving target with $\mathbf{v}_t = 1\text{m/s}$ and $\alpha = \pi/2$. The target's reflectivity is 10% of the reflectivity of the stationary scatterers and its echoes are barely detectable in the original data (see Figure 5-(a)). As we see from the results, i.e., figures 5-(b) and 5-(c) almost perfect separation is achieved between the stationary and the moving targets echoes. There is some noise at the edges of the slow time window in the sparse component but this does not really affect the imaging results.

These observations are further explained in the paper where (i) we provide analysis of the nuclear norms for SAR data tensors corresponding to stationary and moving targets; (ii) we present lower and upper bounds on the values of the tensor nuclear norm, and show how these bounds can be used to explain the observed phenomena; (iii) we prove that the bounds are achieved for limiting edge cases, and that the stationary background and the moving target tend towards those cases.

To summarize, the results of this paper demonstrate that SAR data naturally admit a tensor represen-

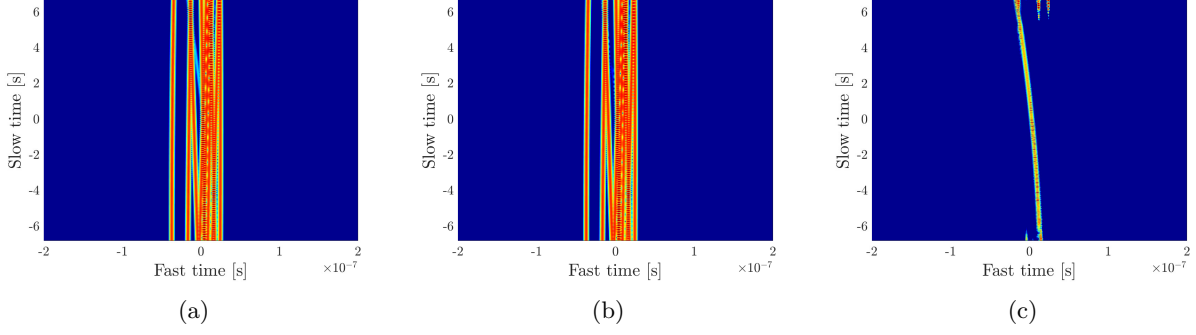


Figure 5: Example of TRPCA applied to SAR data corresponding to a slow moving target. The simulation setup shown in Figure 1 is used with the moving target velocity $\mathbf{v}_t = 1\text{m/s}$ and $\alpha = \pi/2$. The target’s reflectivity is 10% of the reflectivity of the stationary scatterers. (a) Original data matrix. The echoes of the weak and slow moving target are barely discernible in the data; (b) RPCA result: the low rank part corresponding to the stationary background; (c) TRPCA result: the sparse part corresponding to the moving target; We observe that TRPCA achieves good data separation in a very challenging setting.

tation, and provide an example to the added benefit of using the tensor decomposition. In an operational system TRPCA could be combined with other methods such as DPCA or matrix RPCA to detect motion in cases where other methods underperform. Further research will go into other possible extensions of RPCA to tensor form, application of tensor representation to other imaging problems such as iSAR and satellite imaging, and optimization of the algorithm’s parameters for robust performance.

The rest of the paper is structured as follows: In Section 2 we present the TRPCA algorithm for the SAR data problem. In Section 3 we investigate the effect of the hyper-parameters on the performance of TRPCA, both numerically, and through analysis of idealized cases. We prove bounds on the tensor nuclear norm, and study how these bounds are related to the observed performance. In Section 4 we present numerical examples of data separation and the corresponding imaging results which illustrate the advantages of tensor over matrix RPCA for SAR problems. We end in Section 5 with our conclusions.

2 TRPCA for SAR data

In this section we describe the TRPCA algorithm for the SAR data problem, starting with discussion of the tensor nuclear norm.

2.1 TRPCA formulation

Using the representation of the SAR data as the tensor \mathcal{A} defined in (1.9), we want to solve the following tensor RPCA optimization

$$\min_{\mathcal{L}+\mathcal{S}=\mathcal{A}} \|\mathcal{L}\|_* + \eta\|\mathcal{S}\|_1, \quad (2.11)$$

where \mathcal{L} and \mathcal{S} are third order tensors, and $\|\cdot\|_*$ is some generalization of the nuclear norm to third order tensors. In Appendix B, we review common tensor decomposition methods, and the extension of the nuclear norm to tensor form as a relaxed rank estimate.

A natural definition of the nuclear norm [13] that extends to higher dimensions is

$$\|\mathcal{A}\|_{*,\mathcal{T}} = \inf \left\{ \sum_{i=1}^r |\lambda_i| \mid \mathcal{A} = \sum_{i=1}^r \lambda_i u_i^1 \otimes u_i^2 \otimes \cdots \otimes u_i^d, \|u_i^j\| = 1, r \in \mathbb{N} \right\}. \quad (2.12)$$

Note that the u_i^j do not need to be orthogonal. As outlined in [13], (2.12) is in general an NP hard problem to compute. Therefore we look for alternative, more tractable definitions of the nuclear norm.

The simplest alternative would be to take the matrix panels with respect to a specific dimension and compute the matrix SVD on every panel separately so that, for example,

$$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \quad A^{(\ell)} = A(\cdot, \cdot, \ell) = U^{(\ell)} \Sigma^{(\ell)} V^{(\ell)H}, \quad \ell = 1, \dots, n_3, \quad (2.13)$$

$$\|\mathcal{A}\|_{*,\mathcal{D}} = \sum_{\ell=1}^{n_3} \|A^{(\ell)}\|_* = \sum_{\ell=1}^{n_3} \sum_{j=1}^{r_\ell} \sigma_j^{(\ell)}. \quad (2.14)$$

This definition retains some of the properties of singular values, for example the Hilbert-Schmidt norm must equal the sum of singular values,

$$\|\mathcal{A}\|_F^2 = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} |a_{i_1, i_2, i_3}|^2 = \sum_{i=1}^r \sigma_i^2 = \sum_{\ell=1}^{n_3} \sum_{j=1}^{r_\ell} [\sigma_j^{(\ell)}]^2. \quad (2.15)$$

We later show that (2.14) serves as an upper bound to (2.12). The decomposition in (2.14) is *decoupled* in a sense (hence the superscript \mathcal{D}), since the singular values of different panels are computed independently. Using this definition in the RPCA algorithm, would result in the regular matrix RPCA on every subaperture.

Another possible extension was introduced in [6, 20, 19], and used for RPCA in [40, 26]. This is based on an extension of matrix multiplication to incorporate a circular convolution with respect to the third dimension,

$$C = \mathcal{A} * \mathcal{B}, \quad C^{(\ell)} = \frac{1}{n_3} \sum_{p=0}^{n_3-1} A^{(p)} B^{(\ell-p) \bmod n_3}. \quad (2.16)$$

This is equivalent to representing the third order tensor as a block-circulant matrix, with the tensor multiplication homeomorphic to the regular matrix multiplication

$$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \frac{1}{\sqrt{n_3}} \begin{pmatrix} A^{(0)} & A^{(1)} & \dots & A^{(n_3-1)} \\ A^{(n_3-1)} & A^{(0)} & \dots & A^{(n_3-2)} \\ & & \ddots & \\ A^{(1)} & A^{(2)} & \dots & A^{(0)} \end{pmatrix} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}. \quad (2.17)$$

Block-circulant matrices can be block-diagonalized by a Discrete Fourier Transform (DFT) matrix,

$$\frac{1}{\sqrt{n_3}} \begin{pmatrix} A^{(0)} & A^{(1)} & \dots & A^{(n_3-1)} \\ A^{(n_3-1)} & A^{(0)} & \dots & A^{(n_3-2)} \\ & & \ddots & \\ A^{(1)} & A^{(2)} & \dots & A^{(0)} \end{pmatrix} \xrightarrow{\mathcal{F}} \begin{pmatrix} \hat{A}^{(0)} & & & \\ & \hat{A}^{(1)} & & \\ & & \ddots & \\ & & & \hat{A}^{(n_3-1)} \end{pmatrix}, \quad (2.18)$$

which is equivalent to

$$\hat{\mathcal{A}} = \mathcal{F}_3 \mathcal{A}, \quad \hat{A}^{(k)} = \frac{1}{\sqrt{n_3}} \sum_{\ell=0}^{n_3-1} \omega_{n_3}^{\ell k} A^{(\ell)}, \quad \omega_{n_3} = e^{\frac{i2\pi}{n_3}}, \quad \hat{A}^{(k)} \in \mathbb{R}^{n_1 \times n_2}, k = 0, \dots, n_3 - 1, \quad (2.19)$$

\mathcal{F}_3 denotes here the discrete Fourier transform in the third direction. Thus, another estimate of the nuclear norm is given by

$$\|\mathcal{A}\|_{*,\mathcal{F}} = \sum_{k=0}^{n_3-1} \|\hat{A}^{(k)}\|_*. \quad (2.20)$$

This last definition of the tensor nuclear norm given by (2.20) proves to be very well suited for the SAR motion detection problem. We show this by a performance analysis, carried out in Section 3.

2.2 TRPCA algorithm

We can now recast (2.11) as

$$\min_{\mathcal{L}+\mathcal{S}=\mathcal{A}} \|\mathcal{L}\|_{*,\mathcal{F}} + \eta_{\mathcal{F}} \|\mathcal{S}\|_1 = \min_{\mathcal{L}+\mathcal{S}=\mathcal{A}} \sum_{\ell=0}^{n_3-1} \|\hat{L}^{(\ell)}\|_* + \eta \|S^{(\ell)}\|_1, \quad (2.21)$$

where $L^{(\ell)}$ and $S^{(\ell)}$ are the subapertures of \mathcal{L} and \mathcal{S} , respectively. As for the matrix case [25], we can solve the constrained optimization problem by an augmented Lagrangian

$$L(\mathcal{L}, \mathcal{S}, \mathcal{Y}, \mu) = \|\mathcal{L}\|_{*,\mathcal{F}} + \eta_{\mathcal{F}} \|\mathcal{S}\|_1 + \langle \mathcal{Y}, \mathcal{A} - \mathcal{L} - \mathcal{S} \rangle + \frac{\mu}{2} \|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F^2, \quad (2.22)$$

where the Hilbert-Schmidt norm and inner product are the natural element-wise product extensions of the matrix case. For more details see Appendix (B.10). We can again solve this iteratively, using the Alternating Direction Method of Multipliers (ADMM), noting that the Hilbert-Schmidt penalty term is, by Parseval's theorem, invariant under DFT,

$$\|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F^2 = \sum_{\ell=0}^{n_3-1} \|A^{(\ell)} - L^{(\ell)} - S^{(\ell)}\|_F^2 = \sum_{p=0}^{n_3-1} \|\hat{A}^{(p)} - \hat{L}^{(p)} - \hat{S}^{(p)}\|_F^2. \quad (2.23)$$

Thus, solving by Alternating Direction Method of Multipliers (ADMM), the \mathcal{L}_k minimization step involves singular value thresholding in the Fourier domain, and the \mathcal{S}_k minimization uses element wise thresholding in the real domain. Thresholding is done via the operator Θ_λ

$$\Theta_\lambda(a) = e^{i \arg a} \cdot \max\{|a| - \lambda, 0\}. \quad (2.24)$$

The algorithm is outlined in Algorithm 1, and is a modification of the Inexact Augmented Lagrange Multiplier (ALM) RPCA algorithm introduced in [25].

Convergence is determined by a constraint tolerance, that is $\|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F \leq \tau$. The elements of $\bar{\mathcal{A}}$ can be efficiently computed in parallel, using the FFT algorithm. The computation of \mathcal{L}_{k+1} involves a truncated SVD. Compared with matrix RPCA, there are now n_3 singular value decompositions at each

Algorithm 1 $\mathcal{A} = \mathcal{L} + \mathcal{S}$ Inexact ALM method

```

1: Input: Observation tensor  $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ 
2:  $\mu_0 = \max_{\ell} \|A^{(\ell)}\|_2, \quad \rho = 1.4$ 
3: while not converged do
4:    $\bar{\mathcal{A}} = \mathcal{F}_3(\mathcal{A} - \mathcal{S}_k - \mu_k^{-1} \mathcal{Y}_k)$ 
5:   for  $\ell = 0, \dots, n_3 - 1$  do
6:      $\left[ \bar{U}_{k+1}^{(\ell)}, \bar{\Sigma}_{k+1}^{(\ell)}, \bar{V}_{k+1}^{(\ell)} \right] = \text{svd} \left( \bar{A}^{(\ell)} \right)$ 
7:      $\hat{L}_{k+1}^{(\ell)} = \bar{U}_{k+1}^{(\ell)} \Theta_{\mu_k^{-1} \left[ \bar{\Sigma}_{k+1}^{(\ell)} \right]} \bar{V}_{k+1}^{(\ell)*}$ 
8:   end for
9:    $\mathcal{L}_{k+1} = \mathcal{F}_3^{-1}(\bar{\mathcal{L}}_{k+1})$ 
10:   $\mathcal{S}_{k+1} = \Theta_{\eta \mu_k^{-1}}[\mathcal{A} - \mathcal{L}_{k+1} + \mu_k^{-1} \mathcal{Y}_k]$ 
11:   $\mathcal{Y}_{k+1} = \mathcal{Y}_k + \mu_k (\mathcal{A} - \mathcal{L}_{k+1} - \mathcal{S}_{k+1})$ 
12:   $\mu_{k+1} = \rho \mu_k$ 
13:   $k \rightarrow k + 1$ 
14: end while
15: return  $\mathcal{L}_{k+1}, \mathcal{S}_{k+1}$ 

```

iteration. However, the complexity of truncated SVD reduces for smaller subapertures because the size of the matrices is smaller. The matrices $\hat{L}_{k+1}^{(\ell)}$ for different $\ell = 0, \dots, n_3 - 1$ can be computed in parallel, which also significantly reduces the computational time.

3 TRPCA performance analysis

We follow here the same approach as in [22] in order to analyze the performance of TRPCA. The key idea is that there is a finite range of values for $\eta_{\mathcal{F}}$ in the objective (2.21) that are admissible $\eta_{\min, \mathcal{F}} \leq \eta_{\mathcal{F}} \leq \eta_{\max, \mathcal{F}}$. Indeed, if $\eta_{\mathcal{F}}$ is too large then the nuclear norm term might be small even for moving targets, i.e. $\|\mathcal{S}\|_* \leq \eta \|\mathcal{S}\|_1$. If, on the other hand, $\eta_{\mathcal{F}}$ is too small the ℓ_1 term might be small even for the stationary background $\|\mathcal{L}\|_* \leq \eta \|\mathcal{L}\|_1$. Thus, we can estimate the quantities, $\eta_{\max, \mathcal{F}}$ and $\eta_{\min, \mathcal{F}}$ by

$$\begin{aligned} \eta_{\max, \mathcal{F}} &= \sup_{\mathcal{S} \text{ moving targets}} \frac{\|\mathcal{S}\|_{*, \mathcal{F}}}{\|\mathcal{S}\|_1}, \\ \eta_{\min, \mathcal{F}} &= \inf_{\mathcal{L} \text{ stationary background}} \frac{\|\mathcal{L}\|_{*, \mathcal{F}}}{\|\mathcal{L}\|_1}. \end{aligned} \tag{3.25}$$

The classes of moving target and low rank data structures can be defined in several ways. Following [22], we use our data model, defined in A, and choose representatives of each to use in simulation.

We wish to choose η small enough so that the ℓ_1 term is favorable for moving targets and large enough such that the nuclear term is favorable for the stationary background. We can define an objective which balances both requirements

$$F(\eta) = \frac{\eta_{\min, \mathcal{F}}}{\eta} + \frac{\eta}{\eta_{\max, \mathcal{F}}}, \tag{3.26}$$

with the optimal value

$$\eta_{\mathcal{F}}^* = \sqrt{\eta_{\max, \mathcal{F}} \eta_{\min, \mathcal{F}}}. \tag{3.27}$$

Moreover, we expect that the larger the ratio $\eta_{\max, \mathcal{F}}/\eta_{\min, \mathcal{F}}$ the better the achieved separation, as the objective would have a wider range of admissible η 's.

We would like to use TRPCA under settings that increase this ratio, i.e., get the smallest possible nuclear norm for the stationary background and the largest possible nuclear norm for the moving target. Our objective in this section is to use this ratio so as to determine optimal values for the hyper-parameters, i.e. the subaperture and overlap sizes. To compute $\eta_{\max, \mathcal{F}}$ and $\eta_{\min, \mathcal{F}}$ as defined in (3.25) we would need to consider all possible scenarios of stationary and moving targets. To get first an insight for how the hyper-parameters affect $\eta_{\max, \mathcal{F}}$ and $\eta_{\min, \mathcal{F}}$, we define them for any specific example of a SAR data tensor $\mathcal{A} = \mathcal{A}_L + \mathcal{A}_S$ as

$$\eta_{\max, \mathcal{F}} = \frac{\|\mathcal{A}_S\|_{*, \mathcal{F}}}{\|\mathcal{A}_S\|_1}, \quad \eta_{\min, \mathcal{F}} = \frac{\|\mathcal{A}_L\|_{*, \mathcal{F}}}{\|\mathcal{A}_L\|_1}. \quad (3.28)$$

For reference, we also investigate these quantities for the decoupled case, that is,

$$\eta_{\max, \mathcal{D}} = \frac{\|\mathcal{A}_S\|_{*, \mathcal{D}}}{\|\mathcal{A}_S\|_1}, \quad \eta_{\min, \mathcal{D}} = \frac{\|\mathcal{A}_L\|_{*, \mathcal{D}}}{\|\mathcal{A}_L\|_1}, \quad (3.29)$$

and compare their behavior to (3.28).

In what follows we first consider in Section 3.1 specific SAR data scenarios and observe how the quantities in (3.28) and (3.29) depend on the hyper-parameters and the moving target's trajectory. Observing that the main quantity that determines these ratios is the tensor nuclear norm, we introduce in Section 3.2 tensor norm inequalities that help us analyze the performance of TRPCA for general SAR data.

3.1 TRPCA performance analysis for specific SAR data scenarios

We consider a stationary background with 10 point scatterers and a single point moving target, with $\mathbf{v}_t = 1\text{m/s}$. The moving target's reflectivity is 10% of the reflectivity of the other, stationary targets. We vary the target's trajectory angle α with respect to the horizontal axis in the 2D plane, that is,

$$\vec{\mathbf{v}}_t = \mathbf{v}_t [\cos \alpha, \sin \alpha, 0], \quad (3.30)$$

between $\alpha = 0$ and $\alpha = \pi/2$ using a step size $\Delta\alpha = \pi/16$. A schematic of the simulation setting is given in Figure 1. The parameters of the simulation are as follows: The total aperture size s_{tot} is fixed at 11.5s and the platform is moving at 200m/s in the y direction, so that the effective aperture size is 2,300m. This yields two data matrices D_L and D_S associated with the stationary background and moving target respectively.

We next let the subaperture size s_{sub} take the values $[0.005s_{\text{tot}}, 0.01s_{\text{tot}}, 0.02s_{\text{tot}}, 0.03s_{\text{tot}}, \dots, 0.3s_{\text{tot}}]$. For each subaperture size, we change the overlap ϑ , as a fraction of the subaperture size, to be 0.1, 0.2, \dots , 0.9. For each of these configurations we create the tensor data structures $\mathcal{A}_L, \mathcal{A}_S$, out of D_L, D_S , according to (1.9). The number of subapertures n_3 is determined by the other parameters through the following formula

$$n_3 = 1 + \left\lceil \frac{s_{\text{tot}} - s_{\text{sub}}}{(1 - \vartheta)s_{\text{sub}}} \right\rceil. \quad (3.31)$$

In Figure 6 we illustrate how n_3 varies as a function of the overlap ϑ for different subaperture sizes. For each configuration, we compute the ℓ_1 norm and the nuclear norm, for both the decoupled and the tensor forms

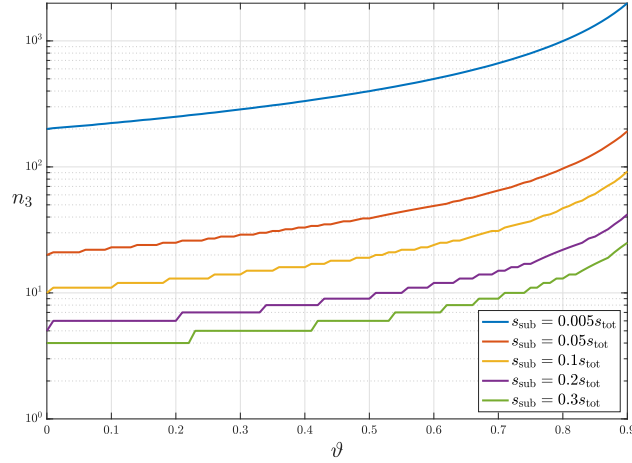


Figure 6: Illustration of how n_3 varies as a function of ϑ in (3.31) for a fixed subaperture size and total aperture. n_3 increases as the size of the subaperture decreases and as ϑ increases.

and plot the ratio of the quantities in (3.29) and (3.28) as function of α and the tensor hyper-parameters in figures 7 and 8 respectively. For the decoupled form (see Figure 7) we observe a weak dependence on the

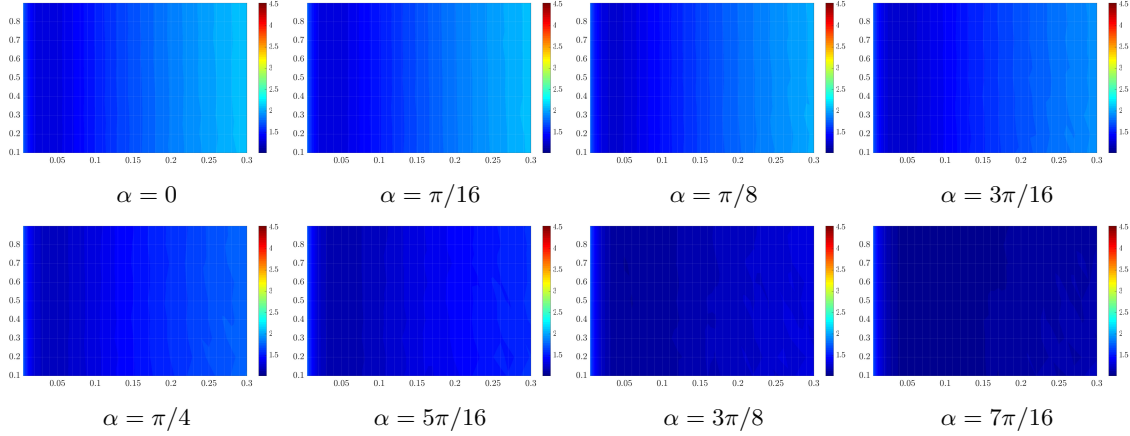


Figure 7: Ratio of decoupled nuclear norm and ℓ_1 norm as a function of data hyper-parameters. We plot $\eta_{\max, \mathcal{D}} / \eta_{\min, \mathcal{D}}$ defined in (3.29), for varying hyper-parameters. The x -axis is the subaperture size and the y -axis is the overlap size. We observe that the ratio is small and favors using larger apertures, with almost no dependence on the subaperture overlap. We can see a slight decrease in value as α increases.

overlap, while the ratio tends to grow with the subaperture size. The maximal value for the ratio is achieved for smaller angles, in consistency with results of regular matrix RPCA.

The tensor case illustrated in Figure 8, presents strong angular dependence, favoring different hyper-parameter configurations at different angles: smaller angles tend to achieve optimal ratio for large apertures with low overlap. However as α increases, the optimal parameter configuration tends towards smaller subapertures with higher overlap. The color scale is the same in figures 7 and 8 indicating that smaller values

are obtained in the decoupled case.

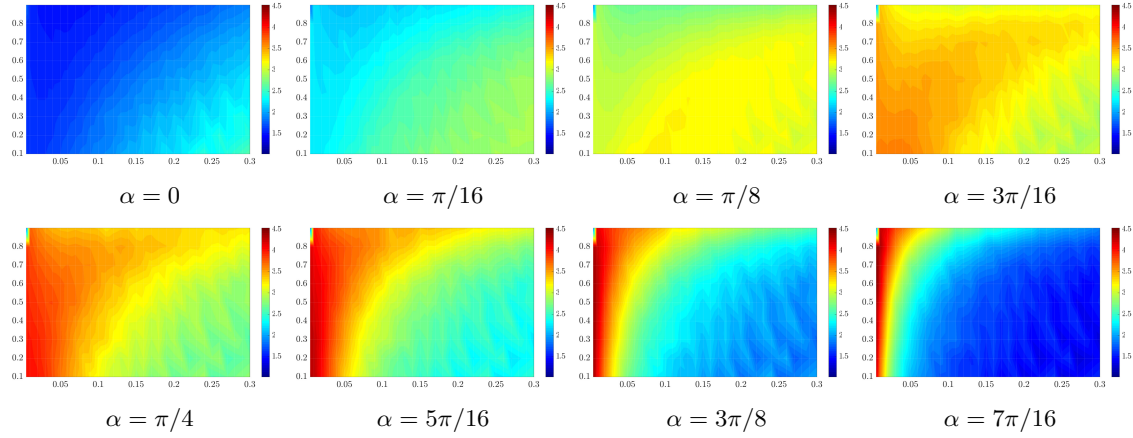


Figure 8: Ratio of tensor nuclear norm and ℓ_1 norm as a function of data hyper-parameters. We plot $\eta_{\max, \mathcal{F}} / \eta_{\min, \mathcal{F}}$, defined in (3.28), for varying hyper-parameters. The x axis and y axis are as in Figure 7. We observe that the dependence on the hyper-parameter changes with the target’s direction. For small angles, larger values are achieved for large subapertures with low overlap. For larger angles, the optimal parameters shift towards smaller subapertures with higher overlap. The most robust behavior is achieved around $\pi/4$. An interpretation of these results is given in Section 3.3

In order to gain more insight, we break down the terms in (3.28), namely the ℓ_1 and nuclear norms of each term, both for the decoupled and tensor forms, and observe their variation in the configuration space. Let us first consider the ratio of the background and moving target’s ℓ_1 norm, as illustrated in Figure 9. We can see that the ℓ_1 norm is insensitive to the direction the target is moving as well as to the hyper-parameters.

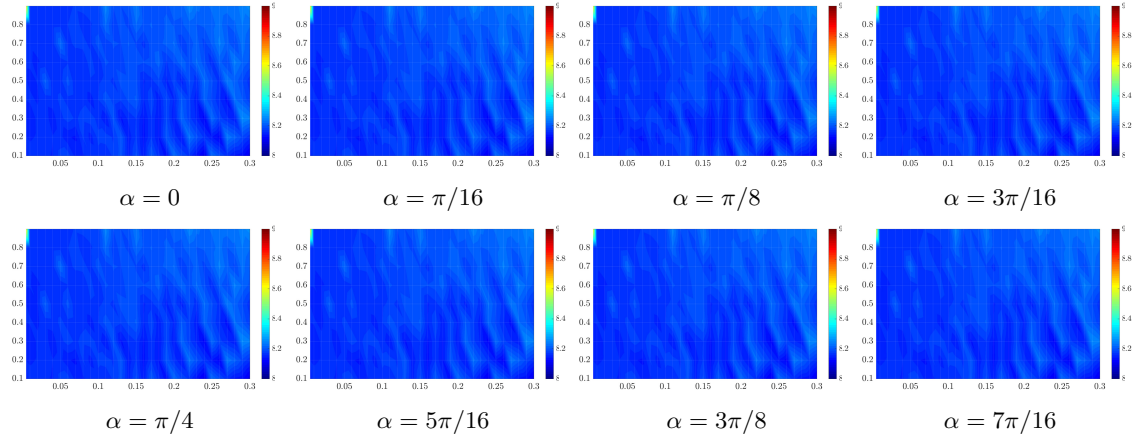


Figure 9: ℓ_1 norm ratio for moving and stationary target as a function of hyper-parameters. We plot the ratio of the background’s and moving target’s ℓ_1 norm for different directions, for varying hyper-parameters. The x axis and y axis are as in Figure 7. As we can see, there is little variation and the ℓ_1 norm values remain similar for all directions.

We next wish to observe the effect of the different parameters on the nuclear norm. To better understand the effect on the tensor nuclear norm, we look at the ratio between the tensor nuclear norm and the decoupled

nuclear norm. Rather than looking at the norms separately, this ratio indicates what is the added benefit of the tensor nuclear norm, factoring out single aperture effects. We plot the ratio of nuclear and decoupled norms for the stationary background as function of the hyper-parameters in Figure 10. Here we observe strong dependence. The tensor nuclear norm is actually smaller than the decoupled norm for small subapertures with high overlap.

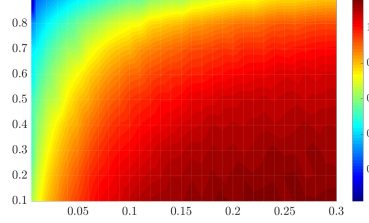


Figure 10: Tensor vs. decoupled nuclear norm for a stationary background, as a function of hyper-parameters. We plot the ratio of tensor and decoupled nuclear norm $\|\mathcal{A}\|_{*,\mathcal{F}}/\|\mathcal{A}\|_{*,\mathcal{D}}$ for a stationary background containing 10 targets, for varying hyper-parameters. The x axis and y axis are as in Figure 7. We see that for every subaperture size, the ratio decreases with the overlap. An interpretation of this result is given in Section 3.3.

We next plot the ratio of the nuclear tensor and decoupled norms for the moving target in Figure 11. In this case we observe strong dependence on both the direction and the hyper-parameters. The highest ratio is achieved for small subapertures with low overlap, but the value, and rate of variation show strong angular dependence.

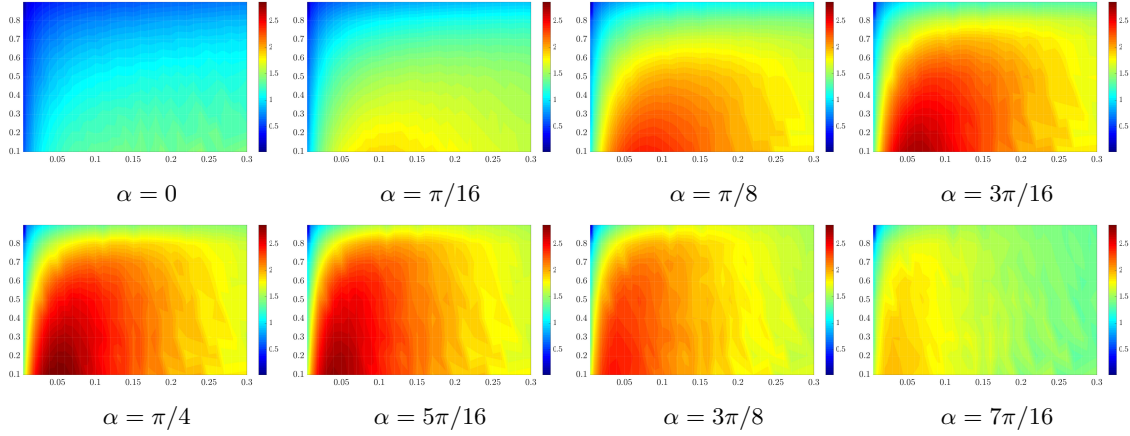


Figure 11: Tensor vs. decoupled nuclear norm for a moving target, as a function of hyper-parameters. We plot the ratio of tensor and decoupled nuclear norm $\|\mathcal{A}\|_{*,\mathcal{F}}/\|\mathcal{A}\|_{*,\mathcal{D}}$ for a moving target for varying hyper-parameters. The x axis and y axis are as in Figure 7. We see that the highest gain is achieved around $\pi/4$, with a low overlap. Smaller angles favor larger subapertures while larger angles favor smaller subapertures. An interpretation of these results is given in Section 3.3.

These results present distinct patterns and raise the following questions:

- Q1. Why, as illustrated in Figure 10 high overlap decreases the ratio of tensor to decoupled nuclear norm for the stationary target?

Q2. What affects the differences in the nuclear norm between moving vs. stationary targets?

Q3. Why, as illustrated in Figure 8, is high η_{\max}/η_{\min} ratio, least sensitive to hyper-parameters, achieved for $\alpha = \pi/4$?

To answer these questions we need to explain how the values of the hyper-parameters and the angle α affect the tensor nuclear norm. This requires a better understanding of the tensor data structure for stationary and moving targets.

To get a better insight on the role of α , we plot in Figure 12 the data traces for a moving target for $\alpha = 0, \pi/4$ and $\pi/2$. The subaperture size and the overlap are fixed to 0.1. In the left column of Figure 12 we plot the the real part of the raw time-domain data for five subapertures while in the second column we plot the real part of the data after performing the Fourier transform along the subaperture index. In the third column we show the row inner product matrix $\hat{A}^{(k)H} \hat{A}^{(k)}$ for a single panel after Fourier.

Our observations are the following: In the raw data, we notice very little variation in the slopes for $\alpha = 0$, $\alpha = \pi/4$ and $\alpha = \pi/2$, i.e., the data look like a translation of the same signal along the different subapertures. However, with a closer look we can see that there is a variation in the phase across the different subapertures which becomes more noticeable as the angle α increases. After performing Fourier, we see that for $\alpha = 0$ the phase is not changing significantly across the different columns of $\hat{A}^{(k)}$. For $\alpha = \pi/4$ the column support of the data is similar to $\alpha = 0$ but more variation in the phase is visible across the different columns of $\hat{A}^{(k)}$. This phase variation translates to the columns of $\hat{A}^{(k)}$ being more orthogonal to each other which results to a diagonal matrix $\hat{A}^{(k)H} \hat{A}^{(k)}$ as seen in the right column. For $\alpha = \pi/2$ the column support is smaller and the phase variation is apparent. As mentioned above the third column of Figure 12 illustrates that the orthogonality between columns of matrix $\hat{A}^{(k)}$ increases as the angle α increases. Indeed, this is manifested by the suppression of the off-diagonal diagonal elements in $\hat{A}^{(k)H} \hat{A}^{(k)}$. We also see that the column range is suppressed for $\pi/2$ (bottom right plot).

These empirical observations require further analysis, provided in the following section where we also answer the questions Q1 to Q3.

3.2 TRPCA analysis for general SAR data

In this section we first give an abstract result for the upper and lower bounds of the tensor nuclear norm. Then we show that the lower bound is obtained for an ideally stationary background with no variation between $A^{(k)}$ for the subapertures while the upper bound is attained for an ideal moving target with exact orthogonality in the columns of $\hat{A}^{(k)}$. We close this section by discussing why the Fourier based tensor nuclear norm is optimal for the SAR data separation problem.

A general result on the bounds of the tensor nuclear norm $\|\cdot\|_{*,\mathcal{F}}$. We have the following Proposition.

Proposition 1. *The tensor nuclear norm, as defined in (2.20), has the following lower and upper bounds*

$$\left(\sum_{i=0}^{n_3-1} \|A^{(i)}\|_*^2 \right)^{1/2} \leq \|\mathcal{A}\|_{*,\mathcal{F}} \leq \sqrt{n_3} \sum_{i=0}^{n_3-1} \|A^{(i)}\|_* \quad (3.32)$$

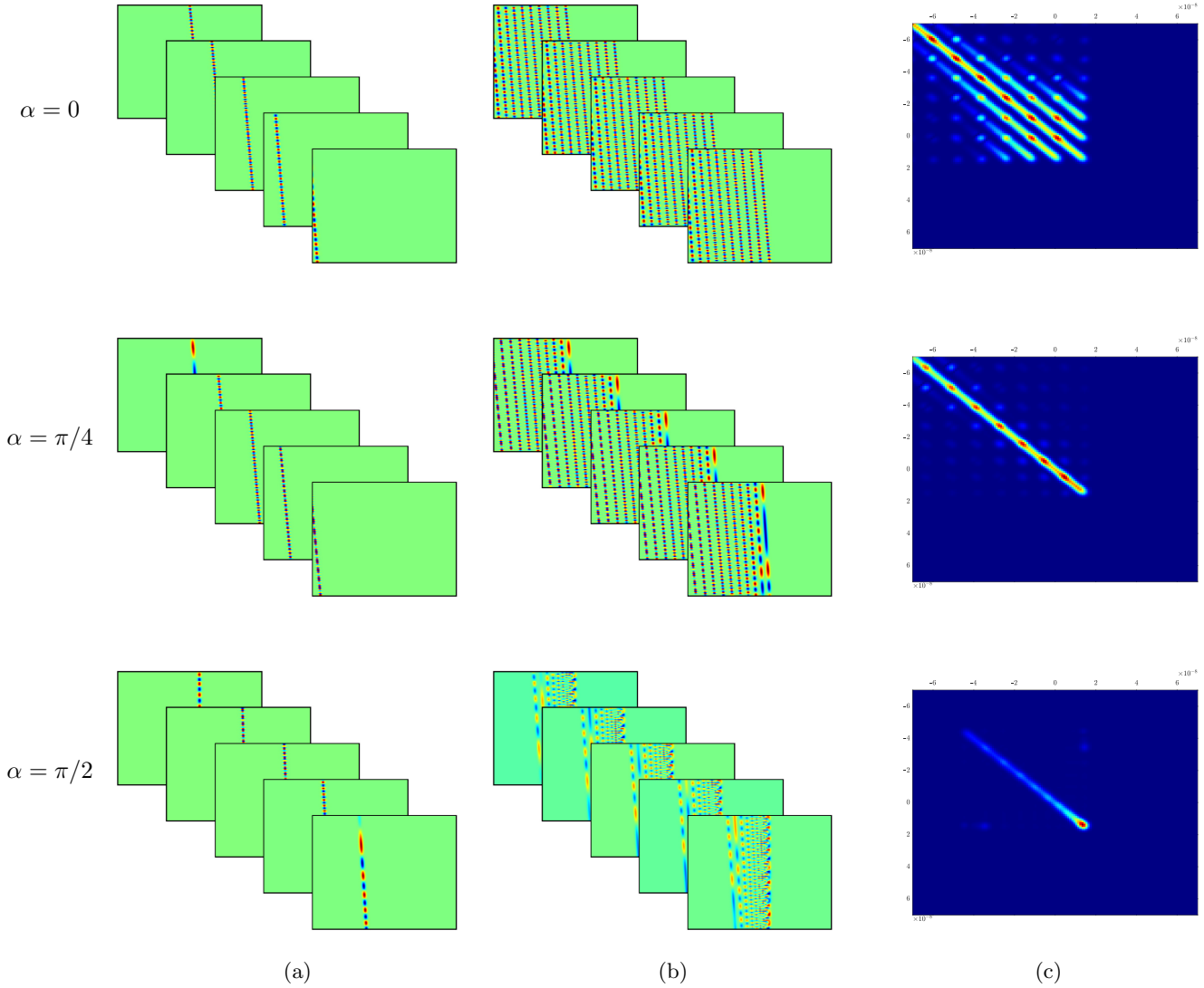


Figure 12: Behavior of SAR data for a moving target with $v_t = 2\text{m/s}$ for different directions α (α as in (3.30); see also Fig. 1); (a) Raw data on five subapertures. Notice very little variation in the slopes between different sub-apertures while phase variation is more noticeable and increases as the angle α increases.; (b) Data after performing Fourier with respect to the subaperture index. Observe that the column support is similar for $\alpha = 0$ and $\alpha = \pi/4$ while it is a lot smaller for $\alpha = \pi/2$. Also notice that the phase variation between the columns of $\hat{A}^{(k)}$ is increasing as the angle increases. (c) $\hat{A}^{(k)H} \hat{A}^{(k)}$ for a single panel after Fourier. For $\alpha = 0$, the column range is large and there are many significant off-block diagonal terms suggesting slow phase variation between the columns of $\hat{A}^{(k)}$. For $\alpha = \pi/4$ the column range is similar to $\alpha = 0$, but off-block diagonal terms are suppressed, suggesting the columns of $\hat{A}^{(k)}$ are closer to being orthogonal. For $\alpha = \pi/2$, off-block diagonal terms are greatly suppressed. Notice also that the total column range is reduced. We give qualitative explanation to the effect each of this observations has on the tensor nuclear norm in Section 3.3.

When all matrices $A^{(i)}$ are the same then

$$\|\mathcal{A}\|_{*,\mathcal{F}} = \left(\sum_{i=0}^{n_3-1} \|A^{(i)}\|_*^2 \right)^{1/2} \quad (3.33)$$

while when the columns of the different matrices $A^{(i)}$ are orthogonal then

$$\|\mathcal{A}\|_{*,\mathcal{F}} = \sqrt{n_3} \sum_{i=0}^{n_3-1} \|A^{(i)}\|_*. \quad (3.34)$$

To prove Proposition 1 we need the following Theorem and Corollary.

Theorem 1. For matrices A_1, A_2, \dots, A_k , $A_i \in \mathbb{C}^{m \times n_i}$, the following holds for the matrix

$$\mathbf{A} = [A_1, A_2, \dots, A_k] \in \mathbb{C}^{m \times N}, \quad N = \sum_{i=1}^k n_i \quad (3.35)$$

$$\left(\sum_{i=1}^k \|A_i\|_*^2 \right)^{1/2} \leq \|\mathbf{A}\|_* \leq \sum_{i=1}^k \|A_i\|_*. \quad (3.36)$$

Proof. The proof is given in Appendix C. □

Corollary 1. For A_1, \dots, A_k in (3.35)

(a) If all the matrices are mutually orthogonal

$$\sum_{i=1}^k \text{rank}(A_i) \leq m, \quad A_i^H A_j = 0, \quad \forall i \neq j, \quad (3.37)$$

the upper bound of (3.36) is attained.

(b) If $A_i = \beta_i A$, $\beta_i \in \mathbb{C}$, then the lower bound of (3.36) is attained and

$$\|\mathbf{A}\|_* = \|\boldsymbol{\beta}\|_2 \|A\|_*, \quad \|\boldsymbol{\beta}\|_2 = \left(\sum_{i=1}^k |\beta_i|^2 \right)^{1/2}. \quad (3.38)$$

Proof. The proof is given in Appendix C. □

We use these general results to prove (3.32).

Proof of (3.32). Write

$$\frac{1}{\sqrt{n_3}} \begin{pmatrix} A^{(0)} & A^{(1)} & \dots & A^{(n_3-1)} \\ A^{(n_3-1)} & A^{(0)} & \dots & A^{(n_3-2)} \\ & & \ddots & \\ A^{(1)} & A^{(2)} & \dots & A^{(0)} \end{pmatrix} = [C_1, C_2, \dots, C_{n_3}], \quad C_i = \frac{1}{\sqrt{n_3}} \begin{pmatrix} A^{(i-1)} \\ A^{(i-2)} \\ \vdots \\ A^{(i)} \end{pmatrix}. \quad (3.39)$$

We now apply Theorem 1 twice. First, for $\mathbf{A} = C_i^T$, $\|A^T\|_* = \|A\|_*$, we get

$$\frac{1}{\sqrt{n_3}} \left(\sum_{i=0}^{n_3-1} \|A^{(i)}\|_*^2 \right)^{1/2} \leq \|C_i\|_* \leq \frac{1}{\sqrt{n_3}} \sum_{i=0}^{n_3-1} \|A^{(i)}\|_*. \quad (3.40)$$

Then, for $\mathbf{A} = [C_1, C_2, \dots, C_{n_3}]$, we get

$$\left(\sum_{i=1}^{n_3} \|C_i\|_*^2 \right)^{1/2} \leq \|\mathcal{A}\|_{*,\mathcal{F}} \leq \sum_{i=1}^{n_3} \|C_i\|_*. \quad (3.41)$$

Plugging the lower and upper bounds of (3.40) in to the lower and upper bound of (3.41), gives the result. \square

We next show how are (3.33) and (3.34) obtained by estimating the nuclear norm $\|\cdot\|_{*,\mathcal{F}}$ for stationary and moving targets. We show that, for high overlap, the $\|\cdot\|_{*,\mathcal{F}}$ norm decreases for stationary targets, while low overlap, increases the norm for moving targets.

$\|\mathcal{A}\|_{*,\mathcal{F}}$ for stationary targets. Let us consider a stationary background, whose returns do not change between pulses. This means that the data traces are the same in every subaperture

$$A^{(i)} = A, \quad i = 0, \dots, n_3 - 1. \quad (3.42)$$

The total decoupled nuclear norm is

$$\|\mathcal{A}\|_{*,\mathcal{D}} = \sum_{i=0}^{n_3-1} \|A^{(i)}\|_* = n_3 \|A\|_*. \quad (3.43)$$

On the other hand for $\|\mathcal{A}\|_{*,\mathcal{F}}$, $\hat{A}^{(i)}$ will have a distinct form since, with respect to the third dimension, we are performing DFT over a constant vector,

$$\hat{A}_{ij}^{(k)} = \frac{1}{\sqrt{n_3}} \sum_{\ell=0}^{n_3-1} \omega_{n_3}^{\ell k} A_{ij}^{(\ell)} = A_{ij} \frac{1}{\sqrt{n_3}} \sum_{\ell=0}^{n_3-1} \omega_{n_3}^{\ell k} = \sqrt{n_3} A_{ij} \delta_{k0}. \quad (3.44)$$

i.e., after performing the DFT in the third dimension, the subapertures have the form

$$\hat{A}^{(k)} = \begin{cases} \sqrt{n_3} A, & k = 0 \\ 0, & k \neq 0 \end{cases}. \quad (3.45)$$

Thus, the total nuclear norm becomes

$$\|\mathcal{A}\|_{*,\mathcal{F}} = \sum_{k=0}^{n_3-1} \|\hat{A}^{(k)}\|_* = \sqrt{n_3} \|A\|_* = \left(\sum_{i=0}^{n_3-1} \|A^{(i)}\|_*^2 \right)^{1/2}. \quad (3.46)$$

This is indeed the lower bound of (3.32). Remark also that since the decoupled nuclear norm is always $\sum_{i=0}^{n_3-1} \|A^{(i)}\|_*$, we get an effective decrease of the nuclear norm by a factor of $\sqrt{n_3}$, with respect to the decoupled norm.

We assumed here a completely stationary background. In reality the echoes from stationary targets do vary along the aperture. However, the variation rate along different subapertures, both in support and in phase, is much slower for stationary data traces. Hence, larger overlap would be beneficial, as it guarantees that when performing DFT, the subapertures are slowly varying, thus suppressing possible amplitude variation and phase decoherence, that will *smear out* the energy over multiple panels after performing DFT. This analysis answers Q1 and is in agreement with the results in Figure 10 where we observed that as the overlap of the subapertures increases, $\|\mathcal{A}\|_{*,\mathcal{F}}/\|\mathcal{A}\|_{*,\mathcal{D}}$ decreases.

$\|\mathcal{A}\|_{*,\mathcal{F}}$ for moving targets. To answer Q2 and Q3, we consider the case of a single moving target, whose data traces are approximately linear. We further assume no overlap between the subapertures, so that there is no column support overlap between different subapertures. This is a good approximation for targets moving linearly in parallel to the projection of $\vec{\mathbf{r}}(0) - \vec{\rho}_o$ on the 2D plane ($\alpha = 0$, see Figure 3a). In this case, every panel would be a translation of the same matrix, without any column overlap. Therefore, we can write

$$A^{(\ell)} = \begin{bmatrix} 0, \dots, 0, A, 0 \dots, 0 \\ \underbrace{\hspace{1.5cm}}_{\ell-1 \text{ times}} \end{bmatrix} \quad (3.47)$$

i.e., the subapertures are copies of the same matrix, supported on disjoint subsets of columns at every subaperture. After performing DFT, the k th panel would be

$$\hat{A}^{(k)} = \frac{1}{\sqrt{n_3}} \left[A, \omega_{n_3}^k A, \omega_{n_3}^{2k} A, \dots, \omega_{n_3}^{(n_3-1)k} A \right], \quad \omega_{n_3} = e^{i2\pi/n_3}. \quad (3.48)$$

Notice that $\hat{A}^{(k)}$ has the same structure as \mathbf{A} in Corollary 1.(b), with $\beta_j = \frac{1}{\sqrt{n_3}} \omega_{n_3}^{jk}$. Since $\|\beta\|_2 = 1$ we have $\|\hat{A}^{(k)}\|_* = \|A\|_*$, and the values of the tensor nuclear norm and decoupled nuclear norm are the same.

This seems in accordance with the results of Figure 11, where we see very little improvement of the nuclear norm for the tensor over the decoupled norm for most hyper-parameter choices when $\alpha = 0$.

But then, why is the performance improving for $\alpha > 0$? We explain this in the following section.

3.3 Interpretation of the nuclear norm angle dependence in SAR data

As we demonstrated in Figure 3, and show in greater detail in Appendix D, in general the data traces are not linear, and the inner product between different subapertures is small. However, their deviation from linearity greatly depends on the direction in which the target is moving. Targets that are moving at angles different than 0 result in higher variation in the slope of the data traces between different subapertures, which leads

to effective orthogonality between data traces of different subapertures. This leads to an increase in the nuclear norm, achieving the upper bound as in Corollary 1(a)

$$\|\widehat{A}^{(k)}\|_* = \frac{1}{\sqrt{n_3}} \sum_{i=0}^{n_3-1} \|A^{(i)}\|_*. \quad (3.49)$$

This explains why $\alpha \neq 0$ improves the nuclear norm ratio, since when the upper bound is achieved the tensor nuclear norm would see an improvement by a factor of $\sqrt{n_3}$ over the decoupled form, since in the decoupled form

$$\|A\|_{*,\mathcal{D}} = \sum_{i=0}^{n_3-1} \|A^{(i)}\|_*. \quad (3.50)$$

While in the tensor form it will be

$$\|A\|_{*,\mathcal{F}} = \sum_{k=0}^{n_3-1} \|\widehat{A}^{(k)}\|_* = \sum_{k=0}^{n_3-1} \sum_{i=0}^{n_3-1} \frac{1}{\sqrt{n_3}} \|A^{(i)}\|_* = \sqrt{n_3} \sum_{i=0}^{n_3-1} \|A^{(i)}\|_*, \quad (3.51)$$

and indeed the upper bound of (3.32).

However, our numerical results suggest that the improvement that is least sensitive to the choice of the hyper-parameters is around $\pi/4$. To explain this, let us recall that in our analysis, we assumed that after performing DFT, we can partition the data traces originating from different subapertures, i.e., that there is no column support overlap between the data of different subapertures. However, in practice there is always a column support overlap, since the subapertures have non zero overlap. In the case where all data traces have the same linear slope, the subaperture overlap (i.e., the number of overlapping rows between successive subapertures) determines the fixed column overlap, proportional to the number of overlapping rows by the slope. However, as the data traces become more curved, the effective slope varies across different subapertures and might even change sign, leading to an increase in the column overlap, no longer determined by the subaperture overlap. This reduces the possible increase in nuclear norm, as the effective column range does not grow at the same rate with increasing number of subapertures. Thus, there is a trade off between the non-linearity of the traces, which implies orthogonality, and their column range support, leading to the most robust performance of the TRPCA around $\alpha = \pi/4$, as is observed in Figure 11. This is also observed in Figure 12, where we can see that for $\alpha = \pi/2$ the column range is limited and, when looking at the panels of $\widehat{A}^{(k)}$ after performing DFT, there is significant variation in the overlap, compared to $\alpha = 0$ and $\alpha = \pi/4$.

Optimality of $\|\cdot\|_{*,\mathcal{F}}$ for SAR data separation. The performance of RPCA in the SAR context improves when the ratio between the nuclear norm of the background and the moving target increases. The main motivation in seeking a tensor based representation is to enhance the low-rank and sparse/full-rank structure of the background and moving targets respectively.

The results of Corollary 1 are a particular case of a more general result, proven in [24]. That result states that the nuclear norm of a tensor \mathcal{T} , can be bound by the nuclear norm of any regular partition of \mathcal{T} (defined in [24]) $\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_k$ by

$$\left(\sum_{i=1}^k \|\mathcal{T}_i\|_*^2 \right)^{1/2} \leq \|\mathcal{T}\|_{*,\mathcal{T}} \leq \sum_{i=1}^k \|\mathcal{T}_i\|_*. \quad (3.52)$$

Specifically, in our case the general statement applies when we choose the partition of \mathcal{A} to be the decomposition of the data into different subapertures $A^{(i)}$, giving

$$\left(\sum_{i=0}^{n_3-1} \|A^{(i)}\|_*^2 \right)^{1/2} \leq \|\mathcal{A}\|_{*,\mathcal{T}} \leq \sum_{i=0}^{n_3-1} \|A^{(i)}\|_*. \quad (3.53)$$

Comparing (3.53) to (3.32) we note that the Fourier based nuclear norm achieves the lower bound for a completely stationary background, while it achieves a factor of $\sqrt{n_3}$ times the upper bound for non linear targets, moving rapidly. It is in this sense, that the Fourier method is optimal for the SAR data separation problem.

4 Numerical results

In this section we use the same setup as before (cf. section 3.1) to generate synthetic SAR data for different configurations. We then use TRPCA and compare its performance with two other methods: matrix RPCA over the entire data, and decoupled RPCA over the different subapertures. We first show the data separation results in Section 4.1 and then the corresponding images in Section 4.2.

4.1 Data separation results

We consider a stationary background with 10 strong point scatterers and a target moving at a slow velocity of $1m/s$. We construct the data matrix D_L and data tensor \mathcal{A}_L for the background and do the same with D_S and \mathcal{A}_S for the moving target. The hyper-parameters are the same as in the previous section. We vary the target's direction α between 0 , $\pi/4$ and $\pi/2$ and perform RPCA in three forms:

- (a) TRPCA with η set to its optimal value:

$$\eta_{\mathcal{F}}^* = \sqrt{\eta_{\max,\mathcal{F}}\eta_{\min,\mathcal{F}}}, \quad \eta_{\max,\mathcal{F}} = \frac{\|\mathcal{A}_S\|_{*,\mathcal{F}}}{\|\mathcal{A}_S\|_1}, \eta_{\min,\mathcal{F}} = \frac{\|\mathcal{A}_L\|_{*,\mathcal{F}}}{\|\mathcal{A}_L\|_1}. \quad (4.54)$$

- (b) ‘Decoupled’: We perform matrix RPCA one subaperture at a time. Here η is also set to its optimal value following [22]:

$$\eta^{(i)*} = \sqrt{\eta_{\max}^{(i)}\eta_{\min}^{(i)}}, \quad \eta_{\max}^{(i)} = \frac{\|A_S^{(i)}\|_*}{\|A_S^{(i)}\|_1}, \eta_{\min}^{(i)} = \frac{\|A_L^{(i)}\|_*}{\|A_L^{(i)}\|_1}. \quad (4.55)$$

- (c) Matrix RPCA over the entire data matrix with η set to its optimal value:

$$\eta^* = \sqrt{\eta_{\max}\eta_{\min}}, \quad \eta_{\max} = \frac{\|D_S\|_*}{\|D_S\|_1}, \eta_{\min} = \frac{\|D_L\|_*}{\|D_L\|_1} \quad (4.56)$$

For methods (a) and (b), we need to reconstruct the separated data on the entire aperture D_L and D_S from the separated data on the overlapping subapertures. From (1.9), one can reconstruct the original data

matrix $D_r(s, t)$ by

$$D_r(s, t) = \sum_{\ell \in X} \theta_\ell A^{(\ell)}(s - \ell \vartheta s_{\text{sub}}), \quad \sum_{\ell} \theta_\ell = 1, \quad X = \{\ell \text{ s.t. } s - \ell \vartheta s_{\text{sub}} \in [0, s_{\text{sub}}]\}. \quad (4.57)$$

We can use any linear combination of the appropriate data entries in the subapertures provided that the weights θ_ℓ sum up to one. The choice used is the ‘innermost’ aperture (i.e. farthest from the edges), which proves to yield stable results,

$$\begin{aligned} D_L(s, t) &= L^{(\ell^*)}(s - \ell^* \vartheta, t) \\ \ell^* &= \arg \min_{\ell} \ell^2 + (\ell - n_3 + 1)^2, \quad s - \ell \vartheta \in [0, s_{\text{sub}}]. \end{aligned} \quad (4.58)$$

The same is done for D_S .

The separation results are illustrated in Figure 13. We observe that the performance of TRPCA improves as the angle α increases. TRPCA outperforms the other two methods for $\alpha \geq \pi/4$, while for small angles, all methods are struggling.

4.2 Imaging results

We present here the imaging results for the angles of Figure 13 for which a clear separation was possible, i.e., for $\alpha \geq \pi/4$. We do not present imaging results when good data separation is not achieved, since the extraction of motion parameters is challenging and prone to errors. To form an image for the moving target, we need to compensate for the target’s velocity when evaluating the SAR functional of (1.6). We explain below how this is done.

Motion Estimation We assume that the sparse part, after performing TRPCA, is composed of a single target. Hence, we can extract $y(s) = \Delta \tau(s)$ from stable peak locations, and compare it to a candidate one, depending on trial target position and velocity

$$f_{\vec{\rho}, \vec{v}}(s) = \tau(s, \vec{\rho} + \vec{v}s) - \tau(s, \vec{\rho}_o). \quad (4.59)$$

We can then extract $\vec{\rho}, \vec{v}$, from solving the following minimum loss problem

$$\vec{\rho}^*, \vec{v}^* = \arg \min_{\vec{\rho}, \vec{v}} \sum_s \mathcal{L}_\delta(y(s) - f_{\vec{\rho}, \vec{v}}(s)). \quad (4.60)$$

With \mathcal{L}_δ a Huber loss:

$$\mathcal{L}_\delta(x) = \begin{cases} \frac{1}{2}x^2, & |x| \leq \delta, \\ \delta(|x| - \frac{1}{2}\delta), & |x| > \delta, \end{cases} \quad \text{with } \delta = 10\Delta t. \quad (4.61)$$

We only need the velocity parameters \vec{v} for the SAR functional. We use MATLAB’s `fmincon` routine to solve the optimization problem and obtain a robust estimate. An illustration of the deviation of the extracted trace from the exact one is given in Figure 14, for $\alpha = \pi/4$.

Since the other two methods do not provide good separation results, it is impossible to extract the velocity parameters from the S part of the data. Therefore to image the moving target, we form an exhaustive 4D

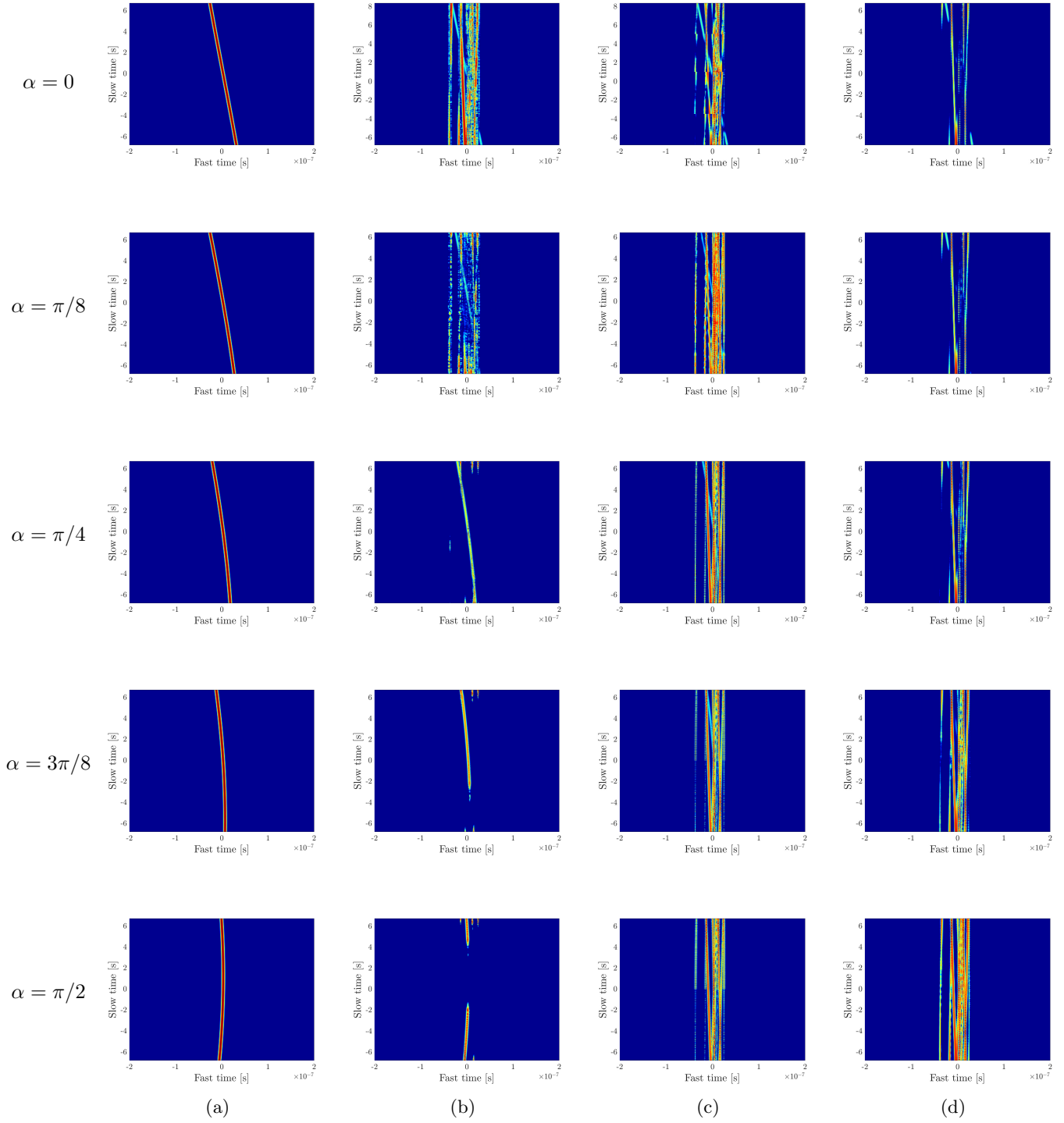


Figure 13: RPCA performance. (a) Original trace of moving target; (b) TRPCA; (c) ‘Decoupled’ RPCA; (d) Matrix RPCA. We can see that for $\alpha \geq \pi/4$ TRPCA outperforms other methods, with its performance improving with larger angles. For small angles, all methods are struggling. Some of the signal is lost around the points where the gradient with respect to the slow time is zero, which are stationary phase points.

imaging functional, where we vary both location and velocity parameters.

$$I^{\text{SAR}}(\vec{\rho}, \vec{v}) = \sum_{j=-n/2}^{n/2} S(s_j, \tau(s_j, \vec{\rho} + s_j \vec{v}) - \tau(s_j, \vec{\rho}_o)). \quad (4.62)$$

We present the result of backpropagation for the other two methods for $I^{\text{SAR}}(\vec{\rho}, \vec{v}_t)$, for $\vec{v} = \vec{v}_t$ the target's actual velocity vector. We present the imaging output of the three RPCA methods in Figure 15. One can see that TRPCA provides a clear image of the moving target, while the other two methods fail to do so.

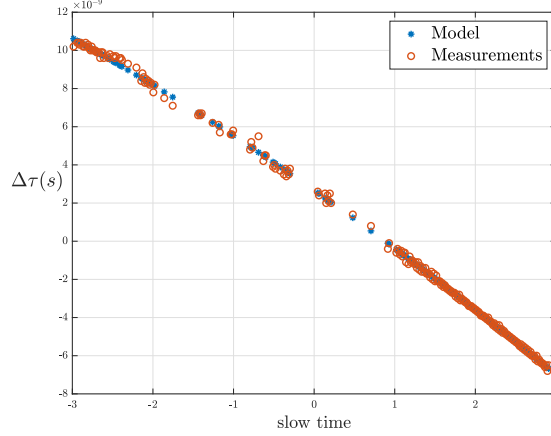


Figure 14: Parameter extraction. Comparison of measured $\Delta\tau(s)$ and analytical one, using the estimated parameters. The sparse part of the data as provided by the TRPCA algorithm is used to determine the velocity parameters, that we subsequently use in the imaging process.

Complex setting TRPCA performs well even when considering more complicated settings. In the following example we have three objects moving with the same velocity of $2m/s$. This can be a model for an extended object whose corners are its strongest reflectors. The background includes 20 scatterers, 10 of which are strong while the other 10 model clutter and have reflectivities comparable to or weaker than the moving targets. We see from the results shown in figures 16 and 17 that TRPCA again performs better than matrix RPCA or the decoupled method.

5 Conclusions

In this paper, we considered the problem of motion detection in Synthetic Aperture Radar (SAR) systems. Our main focus was on separating the data originating from moving targets from those of the stationary background. This separation is necessary since SAR's fundamental assumption is that the objects to be imaged do not vary between successive acquisitions.

We built upon the previously introduced Robust Principal Component Analysis (RPCA) algorithm, separating the data matrix D into low-rank, L , and sparse parts, S , associated with the background and the moving target, respectively. Separation is achieved by solving a convex optimization problem that minimizes

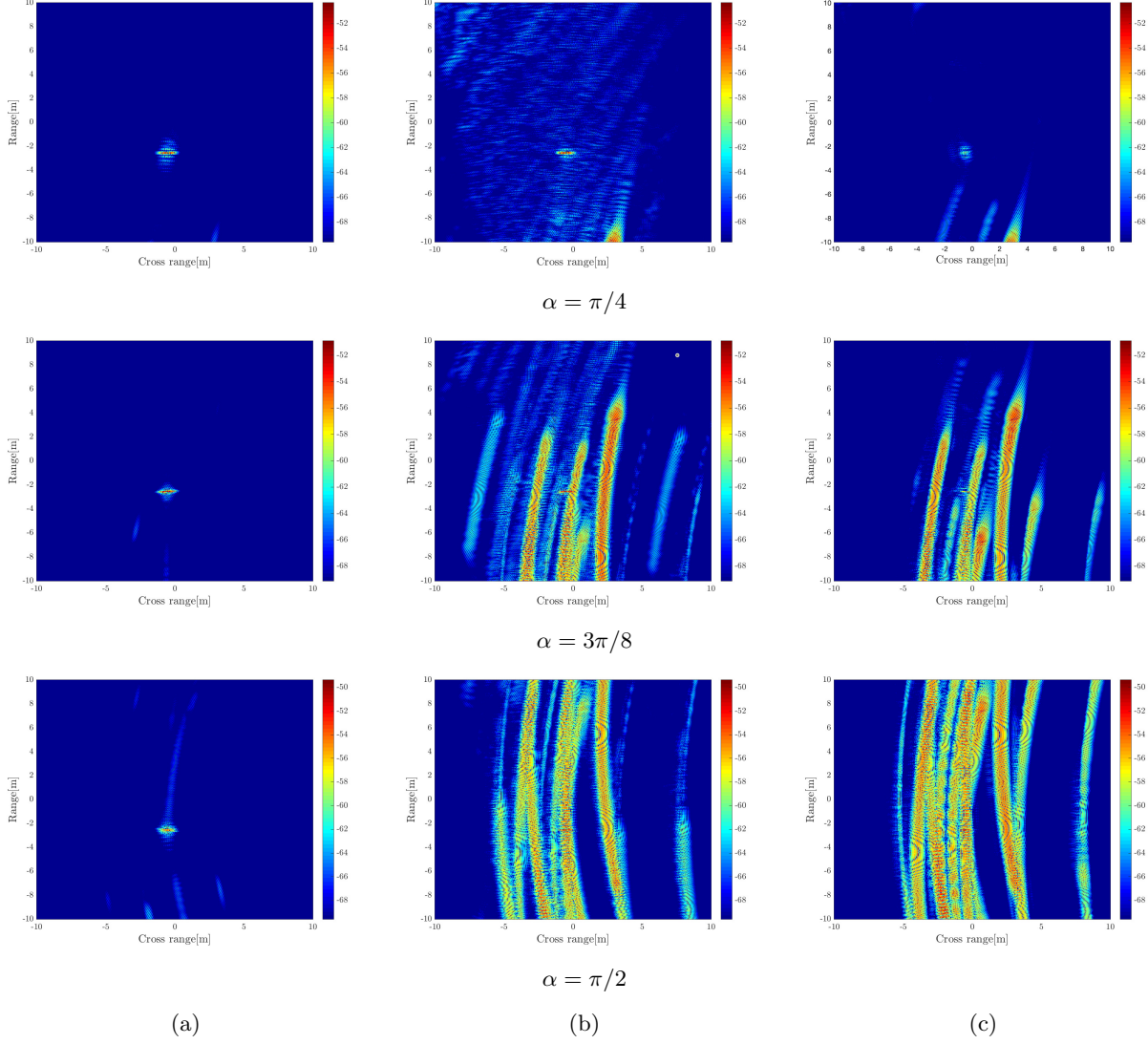


Figure 15: Imaging results for the moving target for $\alpha = \pi/4, 3\pi/8, \pi/2$. (a) using TRPCA ; (b) decoupled RPCA (c) Large aperture matrix RPCA. We observe that TRPCA provides superior imaging results even for large angles ($\alpha = \pi/2$), whereas the other methods fail to suppress the strongly scattering background.

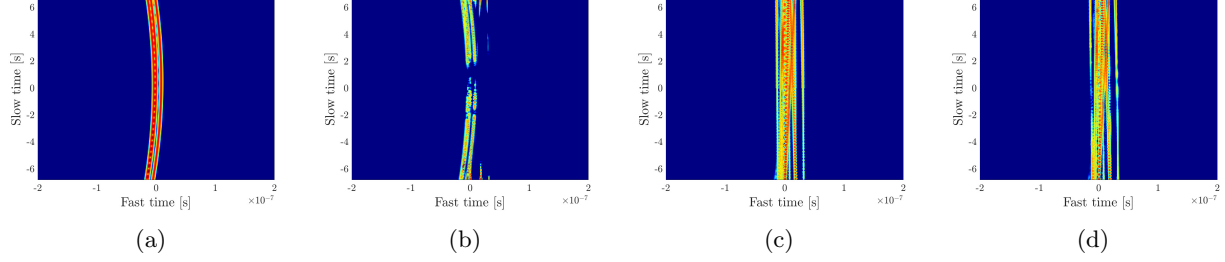


Figure 16: RPCA performance for multiple moving targets, $\alpha = \pi/2$. (a) Original trace of moving targets; (b) TRPCA; (c) 'Decoupled' RPCA; (d) Matrix RPCA. We see that TRPCA again outperforms the other methods, even when there are multiple moving targets and clutter in the background.

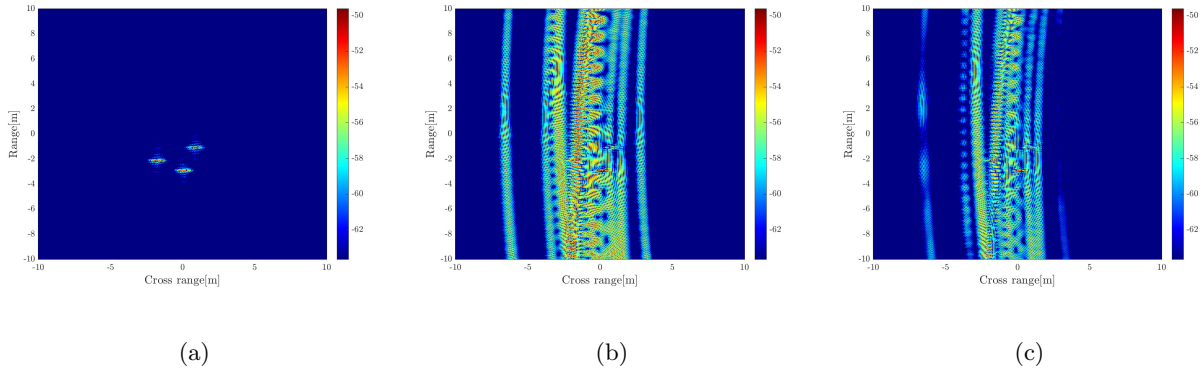


Figure 17: Imaging results for multiple moving targets, $\alpha = \pi/2$. (a) using TRPCA ; (b) 'Decoupled' RPCA (c) Large aperture matrix RPCA. All targets are captured well by TRPCA, while being lost in the background for the other methods.

the objective $\|L\|_* + \eta\|S\|_1$. RPCA has been shown effective in detecting moving targets. However, there are limitations in terms of the speed and the direction of the moving target. In particular, matrix RPCA fails for targets that are slowly moving, or moving in a direction that is parallel to the direction to the SAR platform.

In order to achieve better separation, we have recasted the data matrix as a third order tensor \mathcal{A} , made of partially overlapping subapertures of the original aperture. We employed a specific form for the nuclear norm, using the Fourier transform, to define a tensor version of RPCA. We show with analysis and numerical simulations that this specific form for the tensor nuclear norm, is well suited for the purpose of SAR motion detection, picking on the non linearity of the phase for moving targets at certain directions, while reducing the norm of the stationary background. Numerical simulations in the X-band SAR surveillance regime demonstrated the performance of TRPCA, especially in challenging cases, where motion separation is unattainable using matrix RPCA. Future work will focus on the development of an automated and robust algorithm which adaptively determines the optimal hyper-parameters. This would involve an estimation of the tensor norms for different moving target configurations. Having an integrated algorithm that automatically chooses the hyper-parameters will elevate this method to a signal processing toolbox that can be used with real data.

This work provides a compelling example of a 2D problem that benefits from its reformulation in tensor form. We have shown that there is a natural specific extension of the nuclear norm for the tensor SAR data, which provides optimal results for the motion detection problem, picking up on features such as curvature, which are lost in the matrix representation.

6 Acknowledgements

The work of M. Leibovich and G. Papanicolaou was partially supported by AFOSR FA9550-18-1-0519. The work of C. Tsogka was partially supported by AFOSR FA9550-17-1-0238 and AFOSR FA9550-18-1-0519.

A The SAR data matrix

We describe here our model for the SAR data matrix for a scene comprised of N small point-like targets. We denote σ_i the reflectivity of the i th target and $\vec{\rho}_i(s)$ its location at slow time s . Considering that the SAR platform emits a pulse $f(t)$, the down-ramped data are obtained by convolving the received echoes with $\overline{f(-t)}$. It is therefore as if the antenna emitted the pulse $f_p(t)$ defined as

$$f_p(t) = \int dt' f(t') \overline{f(t' - t)}. \quad (\text{A.1})$$

Considering the free-space causal Green's function for the scalar wave equation

$$G(\vec{r}, t, \vec{r}', t') = \frac{\delta(\|\vec{r} - \vec{r}'\| - c(t - t'))}{4\pi\|\vec{r} - \vec{r}'\|} \mathbb{1}_{t > t'}, \quad (\text{A.2})$$

and assuming

$$\|\vec{\rho}_i - \vec{\rho}_j\| \ll \|\vec{r}(s) - \vec{\rho}_i\|, \forall i, s \quad (\text{A.3})$$

we can model the down-ramped SAR data matrix as

$$D_r(s, t) \propto \sum_{i=1}^N \sigma_i f_p(t - \Delta\tau_i(s)), \quad (\text{A.4})$$

i.e., the data are proportional to a superposition of pulses f_p shifted by $\Delta\tau_i(s)$ and multiplied by σ_i , the reflectivity of each target.

The difference travel time $\Delta\tau_i(s)$ is the round trip travel time from the antenna location $\vec{r}(s)$ to the target $\vec{\rho}_i(s)$ from which the round trip travel time from the antenna location $\vec{r}(s)$ to the reference location $\vec{\rho}_o$ is subtracted to account for the range compression step of down-ramping,

$$\Delta\tau_i(s) = \frac{2(\|\vec{r}(s) - \vec{\rho}_i(s)\| - \|\vec{r}(s) - \vec{\rho}_o\|)}{c}. \quad (\text{A.5})$$

We used here the start-stop approximation which neglects the targets' displacement during the round trip travel time. This is justified in radar because the electromagnetic waves travel at the speed of light, which is many orders of magnitude larger than the speed of the targets and the platform.

It is common to assume that $f_p(t)$ consists of a base-band waveform $f_B(t)$ modulated by a carrier frequency $\nu_o = \omega_o/(2\pi)$,

$$f_p(t) = \cos(\omega_o t) f_B(t). \quad (\text{A.6})$$

Its Fourier transform is

$$\widehat{f}_p(\omega) = \int dt f_p(t) e^{i\omega t} = \frac{1}{2} \left[\widehat{f}_B(\omega + \omega_o) + \widehat{f}_B(\omega - \omega_o) \right]. \quad (\text{A.7})$$

Here $\widehat{f}_B(\omega)$ is supported in the interval $[-\pi B, \pi B]$, where B is the bandwidth while $\widehat{f}(\omega)$ is supported in $[-\omega_o - \pi B, -\omega_o + \pi B] \cup [\omega_o - \pi B, \omega_o + \pi B]$.

The down-ramped SAR data matrix takes the form

$$D_r(s, t) = \sum_{i=1}^N \sigma_i \cos(\omega_o(t - \Delta\tau_i(s))) f_B(t - \Delta\tau_i(s)), \quad (\text{A.8})$$

and its Fourier transform is

$$\widehat{D}_r(s, \omega) = \sum_{i=1}^N \frac{e^{-i\omega\Delta\tau_i(s)}}{2} \left[\widehat{f}_B(\omega + \omega_o) + \widehat{f}_B(\omega - \omega_o) \right]. \quad (\text{A.9})$$

A lossless baseband transformation, introduced in [22], is applied to the data, removing the carrier frequency ω_o

$$\widehat{D}_{rB}(s, \omega) = \widehat{h}_{LP}(\omega) \widehat{D}_o(s, \omega) = e^{-i(\omega - \omega_o)\Delta\tau(s)} \widehat{f}_B(\omega). \quad (\text{A.10})$$

$$\widehat{h}_{LP}(\omega) = \begin{cases} 2 & |\omega| \leq 2B \\ 0 & |\omega| > \alpha B \end{cases}, \quad \alpha B \ll \omega_o. \quad (\text{A.11})$$

A.1 Matrix robust principal component analysis

To produce good imaging results, one needs to detect and separate the echoes corresponding to moving targets from the ones of the complex background.

We follow here the approach proposed in [5], where the robust principal component analysis was used for separating the data matrix \mathcal{D} as in (1.3) in two subsets: the echoes due to stationary targets that form the low rank part of the data matrix and the moving targets echoes which constitute the sparse part. RPCA consists of solving the following convex optimization problem

$$\begin{aligned} \min_{L, S \in \mathbb{C}^{n_1 \times n_2}} \quad & \|L\|_* + \eta \|S\|_1 \\ \text{subject to} \quad & L + S = D. \end{aligned} \quad (\text{A.12})$$

The idea is that the rank of a matrix, which is a non convex objective, can be relaxed to the nuclear norm, that is the sum of singular values, while the element wise ℓ_1 norm, promotes sparsity. The parameter η balances the ratio between the nuclear norm of L and the ℓ_1 -norm of S . The recommended value for η proposed in [7] and used for SAR data in [5] is

$$\eta = \frac{1}{\sqrt{\max\{n_1, n_2\}}}. \quad (\text{A.13})$$

It was shown in [5] that with this choice for η the RPCA algorithm is sensitive to the window size of the data.

As was shown in [22], an optimal choice of η exists for which the separation is robust for the SAR problem. By evaluating the different norms for stationary and moving objects, one can find a range of admissible values

for η , and an optimal one can be derived for a target moving at a velocity v_t ,

$$\eta^* = \sqrt{\frac{\Delta s B \Delta t}{4S\sqrt{\pi}}} \sqrt{\frac{1}{\sqrt{\frac{N(\vec{v}_t)B\Delta t}{2\sqrt{\pi}}} + \frac{1}{2}} \frac{\frac{\sqrt{2}N(\vec{v}_t)B\Delta t}{\pi} + 1}{2}}, \quad (\text{A.14})$$

where, $N(\vec{v}_t)$ is the column support of the moving target, to leading order,

$$N(\vec{v}_t) \approx \frac{4S}{\Delta t} \frac{1}{\|\vec{r}(0) - \vec{\rho}_o(0)\|} (\vec{r}(0) - \vec{\rho}_o(0)) \cdot \frac{v_t}{c}, \quad (\text{A.15})$$

and S is the total slow-time aperture size. In [22], it is also shown, that the separation improves as $N(\vec{v}_t)$ increases.

B Tensor representation and decomposition

Matrix decomposition methods such as SVD, Principal Component Analysis (PCA), Non negative Matrix Factorization (NMF), have enjoyed tremendous success in data science, and have proved to be an essential tool in processing high volume data, from dimensionality reduction to classification.

Tensors are extension of vectors and matrices to higher orders. For example, a p -order tensor \mathcal{A} is defined as

$$\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \cdots \times n_p}, \quad a_{i_1, i_2, \dots, i_p} \in \mathbb{C}, \quad i_k \in \{1, \dots, n_k\} \quad (\text{B.1})$$

The use of tensors is appealing, since it generalizes the properties of matrix data structures, and allows for the representation of more complex patterns.

B.1 Tensor Decomposition methods

Indeed, tensor representations and decomposition of data have been of great interest for several decades. They arise naturally in applications involving high dimensional data, historically in psychometrics [35], chemometrics [1], and more recently in statistics [32], bioinformatics [17], finance [18], signal processing [9] and many more.

A tensor decomposition is defined as a representation of the tensor as a sum of other tensors. For example, a rank-1 SVD decomposition would be of the form

$$a_{i_1, i_2, \dots, i_p} = \sum_{j=1}^r \sigma_j u_{i_1}^{(j,1)} \otimes u_{i_2}^{(j,2)} \cdots \otimes u_{i_p}^{(j,p)}, \quad u^{(j,k)} \in \mathbb{R}^{n_k}, \|u^{(j,k)}\|_2 = 1, \quad (\text{B.2})$$

where $\{u^{(\cdot,k)}\}$ are an orthogonal basis of \mathbb{R}^{n_k} .

We can see that the regular SVD is a particular case of this decomposition

$$a_{i_1, i_2} = \sum_{j=1}^r \sigma_j u_{i_1}^{(j)} v_{i_2}^{(j)*}, \quad (\text{B.3})$$

identifying $u^{(j)} = u^{(j,1)}, v^{(j)} = u^{(j,2)*}$.

However, some of the properties of matrices do not carry on to higher dimensional objects. Decompositions of the form of (B.2) do not exist for general higher order tensors (there are multiple counter examples). Notice that the orthogonality requirement requires $r \leq \min_k n_k$, which does not hold in general. Thus, a need arises for other extensions of matrix decomposition.

Some extensions of matrix decomposition such as Tucker Higher Order Singular Value Decomposition (HOSVD), Canonical Polyadic (CPD), or Tensor Train (TT) have enjoyed applicability [21], but they involve a more complex structure than (B.2). For example, in HOSVD [10], the singular values are replaced by a core tensor so that the representation is

$$a_{i_1, i_2, \dots, i_p} = \sum_{j_1=1}^{r_1} \dots \sum_{j_p=1}^{r_p} \sigma_{\mathbf{J}} u_{i_1}^{(j_1,1)} \otimes u_{i_2}^{(j_2,2)} \dots \otimes u_{i_p}^{(j_p,p)} \quad (\text{B.4})$$

$$u^{(j_k,k)} \in \mathbb{R}^{n_k}, \quad \|u^{(j_k,k)}\|_2 = 1, \quad \mathbf{J} = \{j_1, \dots, j_p\}$$

Specifically, for our purposes, the notion of a rank of tensor, in the sense of a decomposition such as B.2, is not well understood, and its computation has been shown to be NP hard [16].

B.2 Tensor nuclear norm and its estimates

RPCA relies on the nuclear norm as a relaxed rank estimate. Recall that for a matrix A the nuclear norm is defined as

$$\|A\|_* = \sum_{i=1}^r \sigma_i(A) = \max_{\|X\|_\sigma \leq 1} \langle A, X \rangle, \quad (\text{B.5})$$

where

$$\langle A, X \rangle = \text{Tr}(A^H X), \quad (\text{B.6})$$

is the regular matrix inner product and

$$\|X\|_\sigma = \max_{\|u\|_2 \leq 1} \|Xu\|_2 = \max_{\|u\|_2, \|v\|_2 \leq 1} \langle X, u \otimes v \rangle. \quad (\text{B.7})$$

is the matrix spectral norm (or two-norm) [14].

The definition (B.5) is also equivalent to

$$\|A\|_* = \inf \left\{ \sum_{i=1}^r |\lambda_i| \mid A = \sum_{i=1}^r \lambda_i u_i \otimes v_i, \|u_i\| = \|v_i\| = 1, r \in \mathbb{N} \right\}. \quad (\text{B.8})$$

A possible extension of (B.8) to higher order is [13]

$$\|\mathcal{A}\|_{*,\mathcal{T}} = \inf \left\{ \sum_{i=1}^r |\lambda_i| \mid \mathcal{A} = \sum_{i=1}^r \lambda_i u_i^1 \otimes u_i^2 \otimes \dots \otimes u_i^d, \|u_i^j\| = 1, r \in \mathbb{N} \right\}. \quad (\text{B.9})$$

Notice that $u_i^{j_1}, u_i^{j_2}$ need not be orthogonal for $j_1 \neq j_2$.

Definition (B.9) retains many of the favorable properties of the matrix nuclear norm. Specifically, it is the dual norm of the tensor spectral norm.

Define for two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ their inner product as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \dots \sum_{i_d=1}^{n_d} a_{i_1, i_2, \dots, i_d}^* b_{i_1, i_2, \dots, i_d} \quad (\text{B.10})$$

The spectral norm is naturally defined via

$$\|\mathcal{A}\|_\sigma = \max_{\|x_k\|_2 \leq 1, k=1, \dots, d} \langle \mathcal{A}, x_1 \otimes x_2 \otimes \dots \otimes x_d \rangle \quad (\text{B.11})$$

and the nuclear norm is then defined as its dual

$$\|\mathcal{A}\|_{*, \mathcal{T}} = \max_{\|\mathcal{X}\|_\sigma \leq 1} \langle \mathcal{A}, \mathcal{X} \rangle. \quad (\text{B.12})$$

These definitions reduce to the usual matrix definition in two dimensions. However, the computation of both spectral and nuclear norm has been shown to be NP-hard problems in general [13].

Commonly used tensor decompositions do not necessarily provide natural extensions for the nuclear norm, or approximate bounds for it. For example, we cannot take the core tensor of (B.4) as a nuclear norm estimate, since $\sum_{\mathbf{J}} |\sigma_{\mathbf{J}}|$ does not obey the triangle inequality, and hence is not a norm (or a convex objective). We thus look for other extensions of singular values to higher dimensions, presented in the main paper.

C Proof of Theorem 1 and Corollary 1

We will first prove Theorem 1 that we recall next.

Theorem. For matrices A_1, A_2, \dots, A_k , $A_i \in \mathbb{C}^{m \times n_i}$, for the matrix

$$\mathbf{A} = [A_1, A_2, \dots, A_k] \in \mathbb{C}^{m \times N}, \quad N = \sum_{i=1}^k n_i, \quad (\text{C.1})$$

the following inequalities hold

$$\left(\sum_{i=1}^k \|A_i\|_*^2 \right)^{1/2} \leq \|\mathbf{A}\|_* \leq \sum_{i=1}^k \|A_i\|_* \quad (\text{C.2})$$

Proof. Write

$$\mathbf{A} = [A_1, \underbrace{0, \dots, 0}_{k-1 \text{ times}}] + [0, A_2, \underbrace{0, \dots, 0}_{k-2 \text{ times}}] + \dots + [0, \dots, 0, \underbrace{A_k}_{k-1 \text{ times}}], \quad (\text{C.3})$$

and apply the triangle inequality to get the upper bound.

The lower bound is attained following [24]. First, we can bound the spectral norm in the following way: write

$$\mathbf{x}^T = [y_1^T, y_2^T, \dots, y_k^T], \quad \mathbf{x} \in \mathbb{C}^N, y_i \in \mathbb{C}^{n_i} \quad (\text{C.4})$$

$$\|\mathbf{A}\|_\sigma = \max_{\mathbf{x} \in \mathbb{C}^N, \|\mathbf{x}\|_2 \leq 1} \|\mathbf{A}\mathbf{x}\|_2 = \max_{y_i \in \mathbb{C}^{n_i}, \sum_{i=1}^k \|y_i\|_2^2 \leq 1} \left\| \sum_{i=1}^k A_i y_i \right\|_2 \leq \max_{y_i \in \mathbb{C}^{n_i}, \sum_{i=1}^k \|y_i\|_2^2 \leq 1} \sum_{i=1}^k \|A_i y_i\|_2. \quad (\text{C.5})$$

Write $\xi_i = \|y_i\|_2$, $y_i = \xi_i \tilde{y}_i$

Then

$$\|\mathbf{A}\|_\sigma \leq \max_{\sum_{i=1}^k \xi_i^2 \leq 1, \|\tilde{y}_i\| \leq 1} \sum_{i=1}^k \xi_i \|A_i \tilde{y}_i\|_2 = \max_{\sum_{i=1}^k \xi_i^2 \leq 1} \sum_{i=1}^k \xi_i \max_{\|\tilde{y}_i\| \leq 1} \|A_i \tilde{y}_i\|_2 = \max_{\sum_{i=1}^k \xi_i^2 \leq 1} \sum_{i=1}^k \xi_i \|A_i\|_\sigma = \left(\sum_{i=1}^k \|A_i\|_\sigma^2 \right)^{1/2}. \quad (\text{C.6})$$

The last result is achieved by noting that, identifying $u_i = \xi_i$, $v_i = \|A_i\|_\sigma$, we have for $v \in \mathbb{R}_+^k$

$$\max_{u \in \mathbb{R}_+^k, \|u\|_2 \leq 1} \langle u, v \rangle = \left\langle \frac{v}{\|v\|_2}, v \right\rangle = \|v\|_2 = \left(\sum_{i=1}^d v_i^2 \right)^{1/2} \quad (\text{C.7})$$

The nuclear norm is defined as

$$\|\mathbf{A}\|_* = \max_{\|\mathbf{X}\|_\sigma \leq 1} \langle \mathbf{A}, \mathbf{X} \rangle = \max_{\|\mathbf{X}\|_\sigma \leq 1} \sum_{i=1}^k \langle A_i, Y_i \rangle, \quad \mathbf{X} = [Y_1, Y_2, \dots, Y_k] \in \mathbb{C}^{m \times N}, Y_i \in \mathbb{C}^{m \times n_i}. \quad (\text{C.8})$$

From the bound on the spectral norm we have

$$\left\{ \mathbf{X} \in \mathbb{C}^{m \times N} \mid \sum_{i=1}^k \|Y_i\|_\sigma^2 \leq 1 \right\} \subseteq \left\{ \mathbf{X} \in \mathbb{C}^{m \times N} \mid \|\mathbf{X}\|_\sigma \leq 1 \right\}. \quad (\text{C.9})$$

Hence,

$$\|\mathbf{A}\|_* = \max_{\|\mathbf{X}\|_\sigma \leq 1} \sum_{i=1}^k \langle A_i, Y_i \rangle \geq \max_{\sum_{i=1}^k \|Y_i\|_\sigma^2 \leq 1} \sum_{i=1}^k \langle A_i, Y_i \rangle \quad (\text{C.10})$$

Define $y_i = \|Y_i\|_\sigma$, $Y_i = y_i \tilde{Y}_i$.

Then,

$$\|\mathbf{A}\|_* \geq \max_{\sum_{i=1}^k y_i^2 \leq 1, \|\tilde{Y}_i\|_\sigma \leq 1} \sum_{i=1}^k y_i \langle A_i, \tilde{Y}_i \rangle = \max_{\sum_{i=1}^k y_i^2 \leq 1} \sum_{i=1}^k y_i \max_{\|\tilde{Y}_i\|_\sigma \leq 1} \langle A_i, \tilde{Y}_i \rangle = \max_{\sum_{i=1}^k y_i^2 \leq 1} \sum_{i=1}^k y_i \|A_i\|_* = \left(\sum_{i=1}^k \|A_i\|_*^2 \right)^{1/2}. \quad (\text{C.11})$$

□

We next prove Corollary 1 that we recall first.

Corollary. For A_1, \dots, A_k in (3.35)

(a) If all the matrices are mutually orthogonal

$$\sum_{i=1}^k \text{rank}(A_i) \leq m, \quad A_i^H A_j = 0, \quad \forall i \neq j, \quad (\text{C.12})$$

the upper bound of (3.36) is attained.

(b) If $A_i = \beta_i A$, $\beta_i \in \mathbb{C}$, then the lower bound of (3.36) is attained and

$$\|\mathbf{A}\|_* = \|\boldsymbol{\beta}\|_2 \|A\|_*, \quad \|\boldsymbol{\beta}\|_2 = \left(\sum_{i=1}^k |\beta_i|^2 \right)^{1/2}. \quad (\text{C.13})$$

Proof. (a) In this case $\mathbf{A}^H \mathbf{A}$ is block diagonal

$$\mathbf{A}^H \mathbf{A} = \begin{pmatrix} A_1^H A_1 & 0 & \cdots & 0 \\ 0 & A_2^H A_2 & \cdots & 0 \\ & & \ddots & \\ 0 & & & A_k^H A_k \end{pmatrix}, \quad (\text{C.14})$$

and the result is trivial, since

$$\|\mathbf{A}\|_* = \sum_{i=1}^N (\lambda_i(\mathbf{A}^H \mathbf{A}))^{1/2} = \sum_{i=1}^k \sum_{j=1}^{n_i} (\lambda_j(A_i^H A_i))^{1/2} = \sum_{i=1}^k \|A_i\|_*. \quad (\text{C.15})$$

(b) Let us look at $\mathbf{A}^H \mathbf{A}$

$$\mathbf{A}^H \mathbf{A} = \begin{pmatrix} |\beta_1|^2 A^H A & \beta_1^* \beta_2 A^H A & \cdots & \beta_1^* \beta_k A^H A \\ \beta_2^* \beta_1 A^H A & |\beta_2|^2 A^H A & \cdots & \cdots \\ & & \ddots & \\ \beta_k^* \beta_1 A^H A & & & |\beta_k|^2 A^H A \end{pmatrix} = \mathcal{B} \otimes A^H A, \quad (\text{C.16})$$

where $\mathcal{B}_{ij} = \beta_i^* \beta_j$, and \otimes is the tensor product

$$A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{p \times q}, \quad C = B \otimes A = \begin{pmatrix} b_{11}A & b_{12}A & \cdots & b_{1q}A \\ b_{21}A & b_{22}A & \cdots & b_{2q}A \\ \vdots & \vdots & & \vdots \\ b_{p1}A & b_{p2}A & \cdots & b_{pq}A \end{pmatrix} \in \mathbb{R}^{mp \times nq} \quad (\text{C.17})$$

\mathcal{B} has an eigenvalue $\|\boldsymbol{\beta}\|_2^2 = \left(\sum_{i=1}^k |\beta_i|^2 \right)$ with corresponding eigenvector $v_j = \beta_j^*$, and it is easy to see it is an Hermitian, rank one matrix. i.e.,

$$\mathcal{B} = T \Lambda T^H, \Lambda = \text{diag}(\|\boldsymbol{\beta}\|_2^2, 0, \dots, 0). \quad (\text{C.18})$$

Thus, $\mathbf{A}^H \mathbf{A}$ can be block-diagonalized by $T \otimes \mathbb{I}$, to get

$$(T \otimes \mathbb{I})^H A^H A (T \otimes \mathbb{I}) = \begin{pmatrix} \|\boldsymbol{\beta}\|_2^2 A^H A & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}. \quad (\text{C.19})$$

Hence $\sigma_j(\mathbf{A}) = (\lambda_j(\mathbf{A}^H \mathbf{A}))^{1/2} = \|\boldsymbol{\beta}\|_2 \sigma_j(A)$, and $\|\mathbf{A}\|_* = \|\boldsymbol{\beta}\|_2 \|A\|_*$. \square

D Estimate of the the cross terms for two panels

We assume an overall quadratic dependence of $\Delta\tau(s)$ on s ,

$$\Delta\tau(s) = a + bs + cs^2. \quad (\text{D.1})$$

Since the total aperture is decoupled into many, small, subapertures, it is reasonable to approximate $\Delta\tau(s)$ as being linear in each panel, with

$$\Delta\tau_\ell(s) = \Delta\tau_\ell(s + \ell\theta) \approx a_\ell + b_\ell s, \quad a_\ell = a + b\ell\theta + c\ell^2\theta^2, b_\ell = b + 2c\ell\theta. \quad (\text{D.2})$$

The inner product of two columns is

$$\begin{aligned} & \sum_s e^{-\frac{B^2}{2}(t_j - \Delta\tau_\ell(s))^2} e^{i\omega\Delta\tau_\ell(s)} e^{-\frac{B^2}{2}(t_k - \Delta\tau_{\ell'}(s))^2} e^{-i\omega\Delta\tau_{\ell'}(s)} \\ &= e^{-\frac{B^2}{2}((t_j - a_\ell)^2 + (t_k - a_{\ell'})^2)} e^{i\omega(a_\ell - a_{\ell'})} \sum_s e^{-\frac{B^2}{2}(-2(t_j - a_\ell)b_\ell s + b_\ell^2 s^2)} e^{i\omega b_\ell s} e^{-\frac{B^2}{2}(-2(t_k - a_{\ell'})b_{\ell'} s + b_{\ell'}^2 s^2)} e^{-i\omega b_{\ell'} s} \\ &= e^{-\frac{B^2}{2}((t_j - a_\ell)^2 + (t_k - a_{\ell'})^2)} e^{i\omega(a_\ell - a_{\ell'})} \sum_s e^{-\frac{B^2}{2}((-2(t_j - a_\ell)b_\ell - 2(t_k - a_{\ell'})b_{\ell'} + i\frac{2\omega}{B^2}(b_\ell - b_{\ell'}))s + (b_\ell^2 + b_{\ell'}^2)s^2)} \\ &\approx e^{-\frac{B^2}{2}((t_j - a_\ell)^2 + (t_k - a_{\ell'})^2)} e^{i\omega(a_\ell - a_{\ell'})} \frac{1}{\Delta s} \int_{s^-}^{s^+} e^{-\frac{B^2}{2}\phi(s)} ds \\ & \phi(s) \equiv (-2(t_j - a_\ell)b_\ell - 2(t_k - a_{\ell'})b_{\ell'} + i\frac{2\omega}{B^2}(b_\ell - b_{\ell'}))s + (b_\ell^2 + b_{\ell'}^2)s^2 \end{aligned} \quad (\text{D.3})$$

For a quadratic form $\phi(s) = \alpha s^2 + \beta s + \gamma$ stationary phase value is at $s^* = -\frac{\beta}{2\alpha}$, and $\phi(s^*) = \gamma - \frac{\beta^2}{4\alpha}$, which implies

$$\begin{aligned} s^* &= \frac{(t_j - a_\ell)b_\ell + (t_k - a_{\ell'})b_{\ell'}}{b_\ell^2 + b_{\ell'}^2} + i\frac{\omega(b_\ell - b_{\ell'})}{B^2(b_\ell^2 + b_{\ell'}^2)}, \\ \phi(s^*) &= -\frac{((t_j - a_\ell)b_\ell + (t_k - a_{\ell'})b_{\ell'} - i\frac{\omega}{B^2}(b_\ell - b_{\ell'}))^2}{b_\ell^2 + b_{\ell'}^2}. \end{aligned} \quad (\text{D.4})$$

Plugging into (D.3), we get for the real part of the argument of the exponent

$$\begin{aligned}
& -\frac{B^2}{2} \left[(t_j - a_\ell)^2 + (t_k - a_{\ell'})^2 - \frac{(t_j - a_\ell)^2 b_\ell^2 - 2(t_j - a_\ell)(t_k - a_{\ell'})b_\ell b_{\ell'} + (t_k - a_{\ell'})^2 b_{\ell'}^2}{b_\ell^2 + b_{\ell'}^2} + \frac{\omega^2(b_\ell - b_{\ell'})^2}{B^4(b_\ell^2 + b_{\ell'}^2)} + i \dots \right] \\
& = -\frac{B^2}{2} \frac{(b_{\ell'}(t_j - a_\ell) - b_\ell(t_k - a_{\ell'}))^2}{b_\ell^2 + b_{\ell'}^2} - \frac{\omega^2}{2B^2} \frac{(b_\ell - b_{\ell'})^2}{b_\ell^2 + b_{\ell'}^2}
\end{aligned} \tag{D.5}$$

We see there is a suppression term, independent of the specific column indices, proportional to the difference in the slopes. Since $\omega \gg B$, a small difference is enough to create a large suppression- effectively making traces with different slopes orthogonal. We also see that a smaller overlap (larger θ , and more distance between the panels improves the results).

References

- [1] C. J. Appellof and E. R. Davidson. Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents. *Analytical Chemistry*, 53(13):2053–2056, 1981.
- [2] S. Barbarossa and A. Scaglione. Autofocusing of SAR images based on the product of high-order ambiguity function. *IEE Proc.-Radar, Sonar Navig.*, 145(5):269–273, 1998.
- [3] M. Boizard, G. Ginolhac, F. Pascal, and P. Forster. Low rank tensor stap filter based on multilinear svd. In *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 305–308. IEEE, 2012.
- [4] L. Borcea, T. Callaghan, and G. Papanicolaou. Motion estimation and imaging of complex scenes with synthetic aperture radar. *Inverse Problems*, 29(5):054011, 2013.
- [5] L. Borcea, T. Callaghan, and G. Papanicolaou. Synthetic aperture radar imaging and motion estimation via robust principal component analysis. *SIAM Journal on Imaging Sciences*, 6(3):1445–1476, 2013.
- [6] K. Braman. Third-order tensors as linear operators on a space of matrices. *Linear Algebra and its Applications*, 433(7):1241–1253, 2010.
- [7] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust Principal Component Analysis? *Journal of ACM*, 58(1):1–37, 2009.
- [8] M. Çetin, I. Stojanovic, O. Onhon, K. Varshney, S. Samadi, W. C. Karl, and A. S. Willsky. Sparsity-driven synthetic aperture radar imaging: Reconstruction, autofocus, moving targets, and compressed sensing. *IEEE Signal Processing Magazine*, 31(4):27–40, July 2014.
- [9] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, 32(2):145–163, 2015.
- [10] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

- [11] J. Ender. Detection and estimation of moving target signals by multi-channel SAR. *AEU International Journal of Electronic Communication*, 50(2):150–156, 1996.
- [12] J. R. Fienup. Detecting Moving Targets in SAR Imagery by Focusing. *IEEE Transactions on Aerospace and Electronic Systems*, 37(3):794–809, 2001.
- [13] S. Friedland and L.-H. Lim. Nuclear norm of higher-order tensors. *Mathematics of Computation*, 87(311):1255–1281, 2018.
- [14] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.
- [15] H. Guo and N. Vaswani. Video denoising via online sparse and low-rank matrix decomposition. In *2016 IEEE Statistical Signal Processing Workshop (SSP)*, pages 1–5. IEEE, 2016.
- [16] C. J. Hillar and L.-H. Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013.
- [17] V. Hore, A. Viñuela, A. Buil, J. Knight, M. I. McCarthy, K. Small, and J. Marchini. Tensor decomposition for multiple-tissue gene expression experiments. *Nature genetics*, 48(9):1094, 2016.
- [18] E. Jondeau, E. Jurczenko, and M. Rockinger. Moment component analysis: An illustration with international stock markets. *Journal of Business & Economic Statistics*, 36(4):576–598, 2018.
- [19] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications*, 34(1):148–172, 2013.
- [20] M. E. Kilmer and C. D. Martin. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3):641–658, 2011.
- [21] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.
- [22] M. Leibovich, G. Papanicolaou, and C. Tsogka. Low rank plus sparse decomposition of synthetic aperture radar data for target imaging. *IEEE Transactions on Computational Imaging*, 2019.
- [23] G. Li, Y.-N. Peng, and X.-G. Xia. Moving target location and imaging using dual-speed velocity SAR. *IET Radar Sonar Navig.*, 1(2):158–163, 2007.
- [24] Z. Li. Bounds on the spectral norm and the nuclear norm of a tensor based on tensor partitions. *SIAM Journal on Matrix Analysis and Applications*, 37(4):1440–1452, 2016.
- [25] Z. Lin, M. Chen, and Y. Ma. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. *ArXiv e-prints*, Sept. 2010.
- [26] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5249–5257, 2016.

- [27] M. Moradikia, S. Samadi, and M. Çetin. Joint sar imaging and multi-feature decomposition from 2-d under-sampled data via low-rankness plus sparsity priors. *IEEE Transactions on Computational Imaging*, 5(1):1–16, 2018.
- [28] M. Moradikia, S. Samadi, and M. Çetin. Joint sar imaging and multi-feature decomposition from 2-d under-sampled data via low-rankness plus sparsity priors. *IEEE Transactions on Computational Imaging*, 5(1):1–16, 2019.
- [29] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1):6–43, 2013.
- [30] C. E. Muehe and M. Labitt. Displaced-phase-center antenna technique. *Lincoln Laboratory Journal*, 12(2):281–296, 2000.
- [31] K.-Y. Ni and S. Rao. Sar moving target imaging using sparse and low-rank decomposition. In *Radar Sensor Technology XVIII*, volume 9077, page 90771D. SPIE, 2014.
- [32] A. Shashua and T. Hazan. Non-negative tensor factorization with applications to statistics and computer vision. In *Proceedings of the 22nd international conference on Machine learning*, pages 792–799. ACM, 2005.
- [33] A. Soğanlı and M. Çetin. Low-rank sparse matrix decomposition for sparsity-driven sar image reconstruction. In *2015 3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, pages 239–243. IEEE, 2015.
- [34] G. W. Stimson. *Introduction to airborne radar 2nd edition*. Scitech Publishing, Inc, 1998.
- [35] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.
- [36] G. Wang, X. Xia, and V. Chen. Dual-Speed SAR Imaging of Moving Targets. *IEEE Transactions on Aerospace and Electronic Systems*, 42(1):368–379, 2006.
- [37] A. E. Waters, A. C. Sankaranarayanan, and R. Baraniuk. Sparcs: Recovering low-rank and sparse matrices from compressive measurements. In *Advances in neural information processing systems*, pages 1089–1097, 2011.
- [38] H. Yan, R. Wang, F. Li, Y. Deng, and Y. Liu. Ground moving target extraction in a multichannel wide-area surveillance sar/gmti system via the relaxed pcp. *IEEE Geoscience and Remote Sensing Letters*, 10(3):617–621, 2012.
- [39] M. Yasin, M. Çetin, and A. S. Khwaja. Sar imaging of moving targets by subaperture based low-rank and sparse decomposition. In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2017.
- [40] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer. Novel methods for multilinear data completion and de-noising based on tensor-svd. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3842–3849, 2014.

- [41] Z. Zhou, X. Li, J. N. Wright, E. J. Candès, and Y. Ma. Stable principal component pursuit. *2010 IEEE International Symposium on Information Theory*, pages 1518–1522, 2010.