Off-the-Grid Low Rank Matrix Recovery and Seismic Data Reconstruction

Oscar López[†], Rajiv Kumar[*], Özgür Yılmaz[†], and Felix Herrmann[*]

[†]Department of Mathematics, University of British Columbia

[*]Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia

Abstract

Matrix sensing problems capitalize on the knowledge that a data matrix of interest exhibits low rank properties. This low dimensional structure often arises because the data matrix is obtained by sampling a smooth function on a regular (or structured) grid. However, in many practical situations the measurements are taken on an irregular grid (that is completely known). This results in an "unstructured data matrix" that is less fit for the low rank model in comparison to its regular counterpart and therefore subject to degraded reconstruction via rank penalization techniques. In this paper we propose and analyze a modified low rank matrix recovery workflow that admits unstructured observations. By incorporating a regularization operator which accurately maps structured data to unstructured data, into the nuclear norm minimization problem, we are able to compensate for data irregularity. Furthermore, by construction our formulation yields output that is supported on a structured grid, so that in effect we also achieve data regularization. We establish recovery error bounds for our methodology and offer matrix sensing and matrix completion numerical experiments including applications to seismic trace interpolation to demonstrate the potential of the approach.

Index Terms

Matrix sensing, matrix completion, nuclear norm relaxation, non-uniform discrete Fourier transform, data regularization, seismic data, seismic trace interpolation.

I. Introduction

The matrix completion problem of reconstructing a low rank matrix $D \in \mathbb{C}^{n\times m}$ given only a few observed entries has received increased amounts of attention due to its extensive applications and implementation success [5], [6]. More generally, the problem of low rank matrix recovery, also known as matrix

sensing, has the goal of recovering a low rank matrix from N linear measurements and is particularly of interest in the case $N \ll nm$ [2], [12]. The key assumption here is that the underlying matrix of interest has few nonzero or quickly decaying singular values, i.e., the matrix exhibits low rank structure.

More precisely, let $D \in \mathbb{C}^{n\times m}$ be a rank r matrix (or a matrix that can be well approximated by a rank r matrix) and suppose that we have the noisy linear measurements

$$b = \mathcal{A}(D) + d \in \mathbb{C}^N,$$

where $\mathcal{A} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^N$ is linear map with $N < nm$ and $d \in \mathbb{C}^N$ represents our measurement noise with $\|d\|_2 \leq \eta$. We compensate for our small number of measurements by assuming that $r \ll \min\{n, m\}$, i.e., we assume that our matix of interest is of low rank relative to the ambient dimensions (or well approximated by such a rank r matrix). With this assumption, perhaps the most popular method to approximate D is to obtain a data estimate $D^\sharp$ as the argument output of the optimization procedure

$$\min_{X\in\mathbb{C}^{n\times m}} \|X\|_* \text{ subject to } \|\mathcal{A}(X) - b\|_2 \leq \eta, \tag{1}$$

where $\|X\|_*$ is the nuclear norm of X and is defined as $\|X\|_* := \sum_{k=1}^{\min(n,m)} \sigma_k(X)$ where $\sigma_k(X)$ denotes the k-th largest singular value of the matrix.

This methodology has been extensively studied with the concept of rank restricted isometry constants playing an important role [2], [12], [15].

Definition 1: For a linear map $\mathcal{A} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^N$ and $r \leq \min\{n, m\}$, the rank restricted isometry constant $\delta_r(\mathcal{A})$ is defined to be the smallest $0 < \delta < 1$ such that

$$(1 - \delta)\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \delta)\|X\|_F^2$$

for all $X \in \mathbb{C}^{n\times m}$ of rank at most r.

This constant provides a measure of the quality of a $\mathcal{A}$ by ensuring that the linear map is nearly orthonormal (if $\delta \ll 1$) when acting on rank r matrices.

It was shown in [15] that if $\delta_{2r}(\mathcal{A}) < \frac{4}{\sqrt{41}}$ and $D^\sharp$ is the solution of (1), then

$$\|D - D^\sharp\|_F \leq \frac{c_1}{\sqrt{r}} \sum_{\ell=r+1}^{\min\{n,m\}} \sigma_\ell(D) + c_2\eta, \tag{2}$$

where $c_1, c_2 > 0$ depend only on $\delta_{2r}(\mathcal{A})$. Notice that the summation term on the right side of the inequality in (2) is the error of best rank r approximation to D. Hence, this summation vanishes if D is rank r. On the other hand, (2) also shows that the approximation algorithm (1) is robust with respect to "model

mismatch", i.e., the error $\|D - D^\sharp\|_F$ is proportional to the error of the best rank r approximation to D and within the noise level. This result is analogous to the compressive sensing framework (see [15]). Furthermore, in [2], nearly isometric linear operators are introduced and shown to have suitable rank restricted isometry constants with high probability. Examples of these operators include ensembles with independent, identically distributed (i.i.d.) subgaussian entries (see [15], [18]).

In the above scenario, the noise model $d \in \mathbb{C}^N$ and $\eta > 0$, can be defined to encompass unstructured (or irregular) measurements, i.e., samples obtained at locations deviated from a predesigned sampling grid. However, in many data collection settings, extra care is taken so that these grid deviations are known to sufficient precision, e.g., via GPS location tracking. Therefore, in these cases, corruption by off-the-grid observations may be considered separately from the noise model and estimation of the noise energy level η since the known deviations add valuable information that can be used to estimate data on the desired structured grid (a process known as data regularization). Throughout this paper, we adopt this noise model that does not include off-the-grid corruption and further assume that the grid deviations are known.

A. Irregular Data

Many science and engineering applications that utilize low rank matrix recovery formulations owe their success to the assumption that the process of data collection obeys some desired structure. In the context of array signal processing [20], partial recordings are arranged in the form of an incomplete data matrix and accurate reconstruction of the unobserved entries hinges on the assumption that a uniform spatial sampling (USS) method is honored. Thus, these type of applications implicitly require that the array of sensors used for data collection adheres to some proposed layout, e.g., equispaced grid. However, this assumption is often unkept in practice due to physical constraints and natural factors so that the desired low rank construction of the data matrix is hindered.

To state such situations rigorously, it is best to first introduce the notations used throughout the paper. We denote complex valued functions of independent real variables using bold letters, e.g., $\mathbf{D} : \mathbb{R}^2 \mapsto \mathbb{C}$. The discretized counterpart matrices are written with capital letters, e.g., $D \in \mathbb{C}^{n\times m}$ where $D(k,\ell)$ denotes the $(k,\ell) \in [n] \times [m]$ entry of the matrix. Typically b, d are used for vectors where the subscript $\ell \in \mathbb{N}$, e.g., b_ℓ, indicates the ℓ-th entry of the vector. We use x, y, z to denote real variables and write the complex exponential as $\mathbf{e}^x := e^{2\pi \mathbb{i} x}$. The superscript, e.g., y^ℓ, are used for sequencing. For a matrix $X \in \mathbb{C}^{n\times m}$, $\sigma_k(X)$ denotes the k-th largest singular value of X, $\|X\|_* := \sum_{k=1}^{\min(n,m)} \sigma_k(X)$ is the nuclear norm of X, and $\|X\|_F := \sqrt{\sum_{k=1}^{n}\sum_{\ell=1}^{m}(X(k,\ell))^2}$ is the Frobenius norm of X. For vectors

$b \in \mathbb{C}^N$, $\|b\|_2 := \sqrt{\sum_{\ell=1}^{N}(b_\ell)^2}$ denotes the ℓ_2 norm. For functions, $\|\mathbf{f}\|_p$ denotes the L^p norm of $\mathbf{f}$, for an operator $\mathcal{A} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^N$ we let $\|\mathcal{A}\|_{F,2} := \sup_{X\in\mathbb{C}^{n\times m},\|X\|_F\leq 1} \|\mathcal{A}(X)\|_2$ denote the Frobenius to ℓ_2 norm and $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$. For any interger $n \in \mathbb{N}$, $[n]$ denotes the set of positive integers up to n: $\{\ell \in \mathbb{N} : 1 \leq \ell \leq n\}$. For matrices, D^* denotes the transpose of the matrix D.

We may now mathematically address our previously described scenario. Let $Q = [-\frac{1}{2}, \frac{1}{2}]^2$ be our domain of interest and $\mathbf{D} : Q \mapsto \mathbb{C}$ be a continuously differentiable function to be sampled on a rectangular equispaced sampling grid. We denote the discretized regular data matrix $D_0 \in \mathbb{C}^{n\times m}$ with underlying structured grid $G_0 = \{(x^k, y^\ell) = (\frac{k-1}{n} - \frac{1}{2}, \frac{\ell-1}{m} - \frac{1}{2}) : k \in [n],\, \ell \in [m]\} \subset Q$ consisting of a discrete collection of equispaced points, so that $D_0(k,\ell) = \mathbf{D}(x^k, y^\ell)$. For a given $0 < \rho < 1$, called the irregularity parameter, the discretized irregular (or unstructured) data matrix is denoted by $D_\rho \in \mathbb{C}^{n\times m}$ with underlying unstructured grid $G_\rho = \{(\tilde{x}^{k,\ell}, \tilde{y}^{k,\ell}) = (\frac{k-1}{n} - \frac{1}{2}, \frac{\ell-1}{m} - \frac{1}{2}) + (\Delta(k,\ell), \Gamma(k,\ell)) : k \in [n],\, \ell \in [m]\} \subset Q$ consisting of a discrete collection of nonequispaced points. The entries of the perturbation matrices $\Delta, \Gamma \in \mathbb{R}^{n\times m}$ define the pointwise deviations of G_ρ from G_0 with restrictions $0 < |\Delta(k,\ell)| \leq \rho\frac{1}{n}$ and $0 < |\Gamma(k,\ell)| \leq \rho\frac{1}{m}$. Therefore $D_\rho(k,\ell) = \mathbf{D}(\tilde{x}^{k,\ell}, \tilde{y}^{k,\ell})$ and notice that the irregularity parameter ρ controls the allowed amount of deviation from the structured grid which is assumed to be bounded by the grid step size in each axis. Finally let L be the Lipschitz constant of $\mathbf{D}$ in Q.

With this notation, a practitioner must often deal with partial recordings of D_ρ and knowledge of G_ρ, for some $\rho > 0$, rather than the desired ensemble D_0. In applications that agree with our data assumptions, the structured data matrix D_0 is more suitable than its unstructured doppelganger D_ρ for the low rank model. In other words, empirically we observe that the error of the best rank r approximation to D_0 is smaller than that of D_ρ when the underlying sampled function $\mathbf{D}$ is continuously differentiable. Notice that this statement does not hold in general since counterexamples can be constructed by a clever choice of Δ, Γ and $\mathbf{D}$. We nonetheless consider such counterexamples as rare exceptions and adopt our empirical observations by assuming $\text{rank}(D_0) \ll \text{rank}(D_\rho)$ or in the robust case $\|D_0\|_* < \|D_\rho\|_*$ or both whenever $\rho > 0$. We see from (2) that such assumptions imply degraded recovery when dealing with unstructured measurements since the error of reconstruction is proportional to the low rank model mismatch. Our experiments will reveal that this is the case experimentally, where the reconstruction degrades as the irregularity parameter, ρ, is allowed to increase.

Our prime motivation comes from low rank matrix recovery techniques that have recently been introduced for seismic data processing. In this context, seismic data can achieve low rank structure when reorganized to an appropriate transform domain. Working with seismic data in such transform domains allows for missing trace interpolation and simultaneous source separation [1], [8], [17], [19], two important

pre-imaging steps. Furthermore, due to the fallibile physical conditions involved in seismic surveys, source and receiver locations are often placed off a pre designed sampling grid. For this reason data regularization, which aims to approximate data on a chosen equispaced grid, is a subsequent independent processing step required to avoid unwanted artifacts in imaging techniques. Thus, seismic data processing falls under the scope of this paper since low rank assumptions are imposed on data sampled on an irregular grid and further notice that additional processing is needed for data regularization. Progress to incorporate off-the-grid seismic data in compressive sensing techniques has been made by [4], [10], [14] without recovery guarantees. This paper extends the work of [14] and provides recovery error bounds for the method.

B. Contributions

We extend the practical reach of matrix sensing by developing and analyzing a procedure that handles unstructured measurements and simultaneously performs data regularization in cases where the deviated sensor locations are known.

We use the known sensor locations to construct a regularization operator $\mathcal{S} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^{n\times m}$ that maps structured data to the corresponding unstructured data and propose to obtain our approximation $D^\sharp$ by solving the following modified version of (1):

$$\min_{X\in\mathbb{C}^{n\times m}} \|X\|_* \text{ subject to } \|\mathcal{A}\mathcal{S}(X) - b\|_2 \leq \tilde{\eta}. \tag{3}$$

Recall $\mathcal{A} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^N$ is the desired linear measurement map and $\tilde{\eta} > 0$ is a program parameter to be specified momentarily. Adopting the notation from Section I-A, our irregular noisy observations of $\mathbf{D}$ are given by $b = \mathcal{A}(D_\rho) + d$ for some $\rho \in (0, 1)$ and noise bound $\|d\|_2 \leq \eta$. Notice that (3) results from (1) by introducing $\mathcal{S}$ in our measurement model and modifying the misfit parameter η to suit the introduction of $\mathcal{S}$.

In this paper we investigate the regularization operator defined by $\mathcal{S} = \mathcal{N}\mathcal{F}^{-1} = \mathcal{N}\mathcal{F}^*$, where $\mathcal{F}$ is a 2D discrete fast Fourier transform (FFT) and $\mathcal{N}$ is a 2D non-uniform discrete fast Fourier transform (NFFT, see [9], [11]) which consists of evaluating a Fourier expansion at nonequispaced locations according to the unstructured grid G_ρ (refered to as a nonuniform discrete Fourier transform of type 2 in [11]).

We now proceed to our main result.

Theorem I.1. *Let $\mathcal{A} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^N$ and $\eta > 0$ be defined as in (1), $\mathcal{S} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^{n\times m}$ be defined as above and D_0 be defined as in Section I-A. Let $D^\sharp$ be the solution of (3) with $\tilde{\eta} = \eta + (\|\mathcal{A}\|_{F,2})\epsilon$, where*

$\epsilon > 0$ is a constant specified in Lemma II.2 that depends on n, m, Δ, Γ and L, the Lipschitz constant of ***D*** *(for a precise value see (7)). If*

$$\delta_{2r}(\mathcal{AS}) < \frac{4}{\sqrt{41}}$$

then we have

$$\|D_0 - D^\sharp\|_F \leq \frac{c_1}{\sqrt{r}} \sum_{\ell=r+1}^{\min\{n,m\}} \sigma_\ell(D_0) + c_2\eta + c_3\epsilon, \tag{4}$$

where $c_1, c_2, c_3 > 0$ are constants depending only on $\delta_{2r}(\mathcal{AS})$ and ϵ depicts the numerical error that is due to the inaccuracies of the regularization operator $\mathcal{S}$.

Our results imply stability of (3) with respect to irregularity and noise subject to the numerical error of $\mathcal{S}$. Thus, the introduction of the regularization operator $\mathcal{S}$ allows us to exploit the optimal decay of singular values resulting from a structured grid while we can still match our irregular observations. Further notice that as a result the output $D^\sharp$ is supported on a structured grid. In the next section, we investigate the construction and numerical error of the operator $\mathcal{S}$ and provide insight on the requirement of $\delta_{2r}(\mathcal{AS})$.

We end by noting that sampling operators from the matrix completion literature do not in general satisfy rank restricted isometry constant requirements (see [6]). Therefore our result above does not apply to the matrix completion case. However, numerical experiments suggest that our proposed approach is empirically effective in the matrix completion setting as well, which we present in Section III-B.

II. Off-the-Grid Matrix Recovery

Using the notation from Section I-A, we introduce a regularization operator $\mathcal{S} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^{n\times m}$ which achieves $\mathcal{S}(D_0) \approx D_\rho$. Thus, $\mathcal{S}$ aims to simulate the effect of the underlying unstructured grid G_ρ, where the symbol $\approx$ is used to imply the anticipated inaccuracy of the regularization operator. To this end we write $D_\rho = \mathcal{S}(D_0) + P$ and use the notation from Theorem I.1 to denote $\epsilon := \|P\|_F$. Therefore, $P \in \mathbb{C}^{n\times m}$ encompasses the difference between the output of our regularization operator and the true irregular matrix and ϵ quantifies the extent of this numerical error.

The manifestation of the ϵ term in our main result implies that the success of (3) will rely on how well our regularization operator performs. It may be helpful to notice in the ideal case where we have perfectly constructed $\mathcal{S}$, then $\epsilon = 0$ and we obtain reconstruction error as we would have expected from observed data on a structured grid subject to the modified rank restricted isometry constant requirement of $\mathcal{AS}$. With this in mind we must consider the accuracy of our regularization operator and how this

additional operator affects the rank restricted isomety constant requirement with respect to the standard low rank matrix recovery setting.

A. Regularization Operator

We assume that the dimensions of our data matrix, n and m, are odd positive integers and define $\tilde{n} = \frac{n-1}{2}$, $\tilde{m} = \frac{m-1}{2}$. While we can expect similar results for general cases, this assumption simplifies our analysis. We apply the centered inverse FFT to our regular data and write:

$$\check{D}_0(u,v) := \mathcal{F}^*(D_0)(u,v) = \\ \frac{1}{nm}\sum_{p=1}^{n}\sum_{q=1}^{m} D_0(p,q)\mathbf{e}^{(u-\tilde{n}-1)x^p}\mathbf{e}^{(v-\tilde{m}-1)y^q}$$

for $u \in [n]$, $v \in [m]$. If the entries of the matrices $I \in [-\tilde{n},\tilde{n}]^{n\times m}$, $J \in [-\tilde{m},\tilde{m}]^{n\times m}$ denote the chosen irregular frequencies to construct $\mathcal{N}$ and we define $\tilde{I} := \frac{1}{n}I$ and $\tilde{J} := \frac{1}{m}J$, then we apply the centered NFFT to $\check{D}_0$ so that we can write the action of our regularization operator on the regular data matrix as

$$\begin{aligned}\mathcal{S}(D_0)(k,\ell) &:= \mathcal{N}(\check{D}_0)(k,\ell) \\ &= \sum_{u=1}^{n}\sum_{v=1}^{m}\check{D}_0(u,v)\mathbf{e}^{-I(k,\ell)(u-\tilde{n}-1)\frac{1}{n}}\mathbf{e}^{-J(k,\ell)(v-\tilde{m}-1)\frac{1}{m}} \\ &= \sum_{u=1}^{n}\sum_{v=1}^{m}\check{D}_0(u,v)\mathbf{e}^{-\tilde{I}(k,\ell)(u-\tilde{n}-1)}\mathbf{e}^{-\tilde{J}(k,\ell)(v-\tilde{m}-1)},\end{aligned}$$

for chosen entries $\tilde{I}(k,\ell), \tilde{J}(k,\ell) \in [-\frac{1}{2},\frac{1}{2}]$, $k \in [n]$, $\ell \in [m]$. Notice that we are defining $\mathcal{N}$ by evaluating a Fourier expansion at nonequispaced locations (NFFT of type 2 in [11]).

Keeping in mind that we wish to construct $\mathcal{S}$ so that $\mathcal{S}(D_0)(k,\ell) \approx D_\rho(k,\ell) = \mathbf{D}(\tilde{x}^{k,\ell},\tilde{y}^{k,\ell})$, we see that we must choose

$$\tilde{I}(k,\ell) = \tilde{x}^{k,\ell} \text{ and } \tilde{J}(k,\ell) = \tilde{y}^{k,\ell} \tag{5}$$

for all $k \in [n]$ and $\ell \in [m]$.

We now turn to investigate the error commited by $\mathcal{S}$.

Theorem II.1. *Let D_0 and D_ρ be defined as in Section I-A and define $\mathcal{S}$ as specified above (see (5)). Then for every $k \in [n]$ and $\ell \in [m]$ we have*

$$|\mathcal{S}(D_0)(k,\ell) - D_\rho(k,\ell)| \leq \\ c_0 + c_0\boldsymbol{g}_n(\Delta(k,\ell))\boldsymbol{g}_m(\Gamma(k,\ell)), \tag{6}$$

where $c_0 = \left(\frac{6L}{\pi(n-1)} + \frac{6L}{\pi(m-1)}\right)$, *$L$ is the Lipschitz constant of $\boldsymbol{D}$ in Q and for $k \in \mathbb{N}$*

$$\boldsymbol{g}_k(\delta) := 2 + \frac{|\sin(\pi k\delta)|}{\pi} \log\left(\frac{\tan\left(\frac{\pi}{2}\left(\frac{k-1}{k} + |\delta|\right)\right)}{\tan\left(\frac{\pi}{2}|\delta|\right)}\right).$$

This result tells us how the regularization operator suffers in accuracy pointwise in terms of our grid deviations, matrix size and the Lipschitz constant of the underlying sampled function **D**.

Notice that if we allow $|\Delta(k,\ell)|, |\Gamma(k,\ell)| \to 0$, the right hand side of (6) does not vanish and therefore does not reflect the true error bound of the perturbation-free case. Instead we obtain an error bound proportional to c_0, analogous to those in [3] where the author considers the error of interpolation via equispaced finite Fourier coefficients relative to the resulting interpolation from the true equispaced Fourier samples in the 1 dimensional case. Generalizing this method of analysis to the 2 dimensional case, we arrive at our result and as an artifact of the proof technique the perturbation-free error bound does not produce the true zero value of this term. Therefore, these results are most meaningful in our assumed setting $|\Delta(k,\ell)|, |\Gamma(k,\ell)| > 0$ as can be seen in Figure 1.

Applying (6) and using the definition of the Frobenius norm, we obtain our desired error estimate:

Lemma II.2. *With the same notations as in Theorem II.1, let $P = D_\rho - \mathcal{S}(D_0)$. Then the error matrix P satisfies*

$$\epsilon := \|P\|_F \leq c_0 \sqrt{\sum_{k=1}^{n}\sum_{\ell=1}^{m}[1 + \boldsymbol{g}_n(\Delta(k,\ell))\boldsymbol{g}_m(\Gamma(k,\ell))]^2}. \tag{7}$$

This lemma gives us our regularization operator bound ϵ in (4).

We end this section by focusing on the term $\delta_r(\mathcal{AS})$. While the condition $\delta_{2r}(\mathcal{AS}) < \frac{4}{\sqrt{41}}$ of our main result seems more restrictive that the condition in (2), the following result serves to show that this assumption holds for many standard measurement operators, $\mathcal{A}$.

Theorem II.3. *Let $\mathcal{A} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^N$ be a normalized subgaussian measurement map, i.e., all the entries*

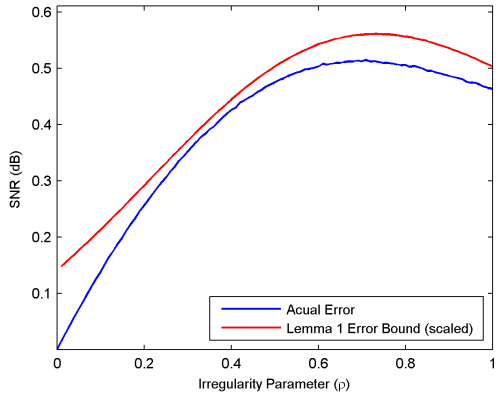

Fig. 1. Plot of irregularity parameter ρ versus mean regularization operator error and scaled regularization operator error bound given by Lemma II.2 with the dataset and grid deviations Δ, Γ introduced in Section III (see Figure 2). The mean is taken over 50 experiments, each with independent Δ, Γ realizations.

$\mathcal{A}(p,k,\ell)$ in the representation

$$\mathcal{A}(D)_p = \frac{1}{\sqrt{N}} \sum_{k=1}^{n} \sum_{\ell=1}^{m} \mathcal{A}(p,k,\ell) D(k,\ell)$$

are independent zero-mean subgaussian random variables with unit variance and subgaussian norm bounded by $\max_{p,k,\ell} \|\mathcal{A}(p,k,\ell)\|_{\psi_2} = K$ with $K \geq \frac{c}{\sqrt{2}}$ for some absolute constant $c > 0$ (here, we have adopted notation from [18]). Define $\mathcal{S}$ as in Theorem II.1 with the entries of each of the perturbation matrices, $\Delta, \Gamma \in \mathbb{R}^{n \times m}$, i.i.d. with any distribution.

Then for any $0 < \delta \leq 1$, $0 < \gamma < \frac{1}{2}$ and rank r matrix, $D \in \mathbb{C}^{n \times m}$,

$$\left(1 - \frac{\delta}{1-2\gamma}\right) \|D\|_F^2 \leq \|\mathcal{A}\mathcal{S}(D)\|_2^2 \leq \left(1 + \frac{\delta}{1-2\gamma}\right) \|D\|_F^2$$

with probability at least $1 - \tau$ provided that

$$N \geq \frac{4K^4}{\bar{c}\delta^2} \left(r(m+n+1) \log\left(1 + \frac{6}{\gamma}\right) + \log(2\tau^{-1}) \right),$$

where $\bar{c} > 0$ is an absolute constant.

In other words, if $\mathcal{A}$ is a subgaussian measurement ensemble and the grid deviations are i.i.d., the proposed operator $\mathcal{A}\mathcal{S}$ inherits its rank restricted isometry constant from $\mathcal{A}$. Therefore, the effect of the regularization operator is negligible in such cases and our methodology is applicable in many standard results of low rank matrix recovery where as above we require $N \propto r(n+m)$ measurements (see [2], [12], [15]). The computational aspects of $\mathcal{S}$ are discussed in the next section.

III. Numerical Experiments

Our previous analysis shows that our proposed implementation (3) remains stable under perturbed measurements according to the behaviour of (7). We now conduct matrix sensing and matrix completion numerical experiments to illustrate that this is also observed in practice. While our analysis is only valid for $\rho \in (0, 1)$, we allow ourselves the extreme values in this section to demonstrate that experimentally our methodology is applicable in these cases.

Datasets: For synthetic matrix sensing and matrix completion experiments, we consider $\mathbf{D} : Q \mapsto \mathbb{R}$ consisting of a sum of six Gaussian functions. Given a center, $(\tilde{x}, \tilde{y}) \in Q$, and two constants, $d_1, d_2 > 0$, each two dimensional Gaussian can be written as $\mathbf{f}(x, y) := e^{-d_1(x-\tilde{x})^2 - d_2(y-\tilde{y})^2} = e^{-d_1(x-\tilde{x})^2} e^{-d_2(y-\tilde{y})^2} := \mathbf{f}_1(x)\mathbf{f}_2(y)$. Thus by obtaining samples on the equispaced grid G_0 each two dimensional Gaussian can be discretized as a matrix via the outer product of two vectors, each resulting from a discretized one dimensional Gaussian function: $\mathbf{f}_1(x)$ and $\mathbf{f}_2(y)$. In other words, each regularly sampled two dimensional Gaussian results in a discretized rank 1 matrix so that $D_0 \in \mathbb{R}^{100\times 100}$ consists of a sum of six rank 1 matrices. For the function we consider here (see Figure 2), we have chosen six distinct function centers that do not agree with each other in either coordinate so that D_0 results in a rank 6 matrix. For $\rho > 0$, the matrix D_ρ generated by sampling $\mathbf{D}$ at the grid points G_ρ is in general full rank (see Figure 2).

To construct the unstructured grid G_ρ for given irregularity parameter ρ, we define our perturbations by sampling from a uniform distribution, so that each $|\Delta(k, \ell)|, |\Gamma(k, \ell)|$ is a uniformly distributed random number between 0 and $\rho\frac{1}{n}, \rho\frac{1}{m}$ respectively and $\operatorname{sgn}(\Delta(k, \ell))$, $\operatorname{sgn}(\Gamma(k, \ell)) \in \{-1, 1\}$ with equal probability for all $(k, \ell) \in [n] \times [m]$ independently. Thus our unstructured data matrix $D_\rho \in \mathbb{R}^{100\times 100}$ is supported on the randomly perturbed grid G_ρ. The unstructured grid is generated randomly for each independent set of experiments (see Figure 2).

For seismic trace interpolation, we utilize the Gulf of Suez 2D seismic line $\tilde{D}_\rho \in \mathbb{C}^{1024\times 354\times 354}$ from [14], where, as before, the subscript ρ indicates the degree of irregularity. These data volumes represent dicretizations of analog finite-energy wavefields in 3 dimensions including time ($t = 1024$), receiver location ($n = 354$) and source location ($m = 354$). Thus our context remains valid since such data pertains to an underlying continously differentiable function. We perform our experiments in a transform domain known as midpoint-offset (m-h) domain with corresponding linear map $\mathcal{M} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^{n\times 2m-1}$ and inverse $\mathcal{M}^{-1} = \mathcal{M}^*$. This transform domain corresponds to a data reorganization by a barycentric change of coordinates and if we think of our matrix as an image (where each image pixel value corresponds to the value of a matrix entry) then this transformation can be thought of as a 45° image rotation. The resulting

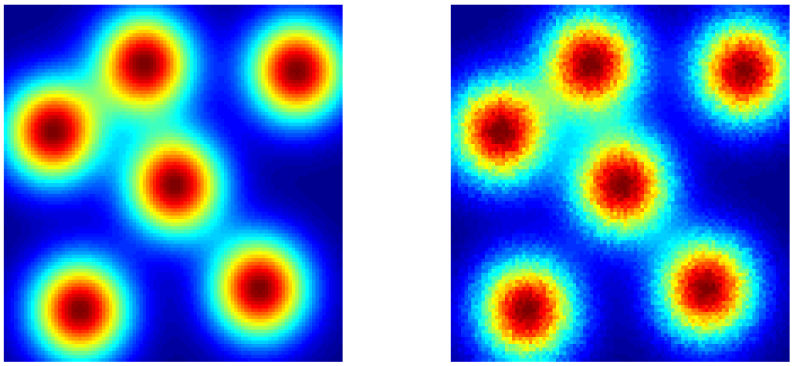

Fig. 2. Synthetic dataset introduced in Section III, which consists of a sum of 6 two-dimensional Gaussian functions. (left) Structured 100×100 data matrix D_0 whose rank is 6. (right) Unstructured 100×100 data matrix D_1 (i.e., irregularity parameter ρ is 1) which is full rank.

matrices in (m-h) domain remain full rank but exhibit quickly decaying singular values in part due to seismic data reciprocity with respect to source and receiver locations (see [1] for details and definitions). To simulate unstructured seismic data ($\tilde{D}_\rho$ for $\rho = \frac{1}{2}$ and $\rho = 1$) we utilize a bicubic spline operator that accurately estimates the irregular grid positions from the densely sampled regular grid positions ($\tilde{D}_0$). This approach acheives a more realistic (but also more computationally expensive) simulation of unstructured seismic data than our chosen regularization operator $\mathcal{S}$ and reduces the inversion crime bias in the implementation of (3), i.e., we do not use the same operator to synthesize and invert data in an inverse problem. For all seismic experiments we randomly perturb the source locations only, so that for each fixed time $t \in [1024]$ we produce the same 354×354 irregular grid G_ρ by constructing Δ (source deviations) as in the previous paragraph while Γ (receiver deviations) is defined as the $n \times m$ zero matrix.

Algorithms: We solve (1) and (3) using LR-BPDN implementation introduced in [1]. This method combines the Pareto curve approach [7] with an SVD-free matrix factorization method in order to handle large-scale low rank matrix recovery problems.

We construct our regularization operator $\mathcal{S} = \mathcal{N}\mathcal{F}^*$ using NFFT 3 software from [9]. Notice that as defined, one can argue that $\mathcal{S}$ takes input data in equispaced Fourier domain and outputs data in nonequispaced Fourier domain according to G_ρ. With this in mind, since our data is in many cases considered to be in space and time domain, it would seem appropriate to instead define $\mathcal{S} = \mathcal{N}^{-1}\mathcal{F}$. However, this choice turns out problematic. Firstly, due to the fact that the operator $\mathcal{N}$ is non-orthogonal

it generally it does not admit an inversion formula. Accordingly, $\mathcal{N}^{-1}$ is expensive to compute. In NFFT 3 software [9], $\mathcal{N}^{-1}$ is approximated by solving a weighted least squares problem, which in turn would affect the computational complexity of implementing (3). For this same reason, our choice of $\mathcal{S} = \mathcal{N}\mathcal{F}^*$ also results in a simplified analysis of the regularization operator error. Because of this, we instead choose to interpret our data in equispaced Fourier domain so that $\mathcal{S} = \mathcal{N}\mathcal{F}^*$ adequately rearranges the time and spatial positions and accurately produces our desired approximated nonequispaced data in Fourier domain.

Despite our previous remarks, it is important to notice that our proposed methodology has a potentially significant added computational cost relative to the original optimization procedure (1). Indeed, the inclusion of the regularization operator in our chosen numerical implementation results in a non-negligible additional complexity at each iteration of LR-BPDN. In this work we have not dedicated much effort to optimize the runtime of (3); however we note that there are options to this end. For example, a parallel version of NFFT 3 software is available in [13] (PNFFT) where the authors demonstrate the high scalability of their algorithm. Such an approach on parallel distributed memory architectures will allow for our methodology to become computationally efficient in large scale data problems.

A. Matrix Sensing

We investigate how the solution of (3) approximates the true solution as we allow an increasing amount of irregularity in our measurements. In accordance with typical compressed sensing experiments, we define $\mathcal{A} : \mathbb{C}^{n\times m} \mapsto \mathbb{C}^{\frac{nm}{2}}$ as a Gaussian measurement ensemble with i.i.d. $\mathcal{N}(0, \frac{2}{nm})$ entries. Such measurement operators are known to achieve suitable rank restricted isometry constants with high probability so that by Theorem II.3 the introduction of the regularization operator does not affect this and allows for our methodology to be successful.

To this end, we fix a sampling operator $\mathcal{A}$. We then vary $\rho \in [0, 1]$ and for each value of ρ we generate our random grid deviation matrices, $\Delta, \Gamma \in \mathbb{C}^{n\times m}$, and implement (1) and (3) with the fixed $\mathcal{A}$. We then repeat this experiment with 500 different randomly generated $\mathcal{A}$ and report the average SNRs for (1) and (3). We repeat the above procedure for noisy Gaussian measurement experiments with noise upper bound $\|d\|_\infty \leq .025\|\mathbf{D}\|_\infty$. Results are shown in Figure 3, where the benefit of our proposed methodology is clear in comparison to the recovery via (1).

Notice in both plots of Figure 3 the SNR curve from (3) begins to increase from $\rho = .8$ to $\rho = 1$. These experimental observations are reflected Lemma II.2 and Figure 1, which demonstrate that the error of the regularization operator decreases in this interval. Intuitively, after a certain extent of allowed grid deviation, the regularization operator begins to benefit by using information from the neighboring grid

points.

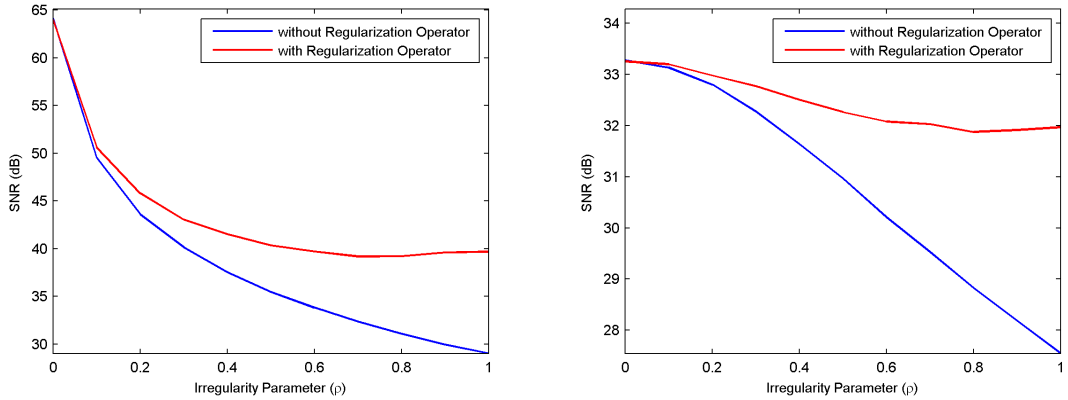

Fig. 3. Results of Gaussian measurement ensemble experiments via LR-BPDN with the dataset introduced in Section III (see Figure 2). In each figure we plot the irregularity parameter ρ versus mean SNR of output of (1) and (3) and report the mean over 500 experiments, each with independent $\mathcal{A}$ realization. (left) Recovery from noiseless Gaussian measurements. (right) Recovery from noisy Gaussian measumrents.

B. Matrix Completion

Although our recovery guarantees do not directly apply to the matrix completion setting, we dedicate this section to demonstrate experimentally that our proposed methodology extends to this scenario. A future direction is to provide recovery guarantees for matrix completion via our methodology where partial matrix entries are sampled on an unstructured grid.

As in the matrix sensing experiments, we allow an increasing amount of irregularity in our measurements. To this end, we fix a set of observed matrix entries $\Omega \subset [n] \times [m]$, constructed by sampling each location independently with various probabilities (including the case $\Omega = [n] \times [m]$, with the goal of data regularization). We then vary $\rho \in [0, 1]$ and for each value of ρ we generate our random grid deviation matrices, $\Delta, \Gamma \in \mathbb{C}^{n \times m}$, and implement (1) and (3) with the fixed Ω. We then repeat this experiment with 500 different randomly generated Ω and report the average SNRs for (1) and (3). We repeat the above procedure for the noisy matrix completion experiments with noise upper bound $\|d\|_\infty \leq .025\|\mathbf{D}\|_\infty$. Results are shown in Figure 4, where our methodology demonstrates improved reconstruction in comparison to (1).

Notice that in the plots of the experiments with 75% missing entries the SNR curves are increasing from $\rho = 0$ to $\rho = .2$ and remain more or less stable at this dB level (this is also apparent to some

extent in the noisy case with 50% missing entries). These observations imply that for highly sub sampled matrix completion experiments, it is beneficial to sample on a perturbed grid and employ our methodology rather than enforce the desired structured grid and reconstruct via (1). We do not understand the reason for these observations and, to the best of our knowledge, have not seen similar observations in the literature when the structure of the unobserved matrix entries is random. We hypothesize that such irregular sampling techniques, along with the randomly generated set of sampled entries Ω, help deal with aliasing uncertainty and thus provide supplementary information that allow $\mathcal{S}$ to act as an enhanced denoising operator. Unfortunately, our current analysis, which involves rank restricted isometry constants, is not suited for the matrix completion scenario so that at the moment we cannot fully investigate this phenomenon and therefore leave this task as future work.

Seismic Trace Interpolation: In seismic field experiments, due to physical or budget constraints several sources (and receivers) are removed which translates to missing traces in otherwise densely sampled data. Since most seismic imaging algorithms require densely sampled data, interpolation to estimate the missing traces is an important preimaging step in order to avoid imaging artifacts caused by subsampling. For this purpose, rank minimization techniques to reconstruct missing traces have been proposed in [1], [19] by recognizing this application as a special instance of the matrix completion problem.

For seismic trace interpolation experiments, we adopt the techniques proposed in [1] with the modification of our proposed methodology (3) and refer the reader to this literature for details and definitions. To explain briefly, we simulate the effect of subsampling by removing 75% of the sources in a jittered random manner which translates to missing columns in each 354×354 time slice. This subsampling technique aims to remove sources randomly while controlling the maximum allowed gap between missing traces. We then apply a Fourier transform along the time axis of our data cube, $\tilde{D}_\rho \in \mathbb{C}^{1024\times 354\times 354}$ (for $\rho = \frac{1}{2}$ and $\rho = 1$) and solve the matrix completion problem for each monochromatic 354×354 frequency slice in a sequential manner in (m-h) domain. As mentioned before, if we consider our matrix as an image, then the (m-h) domain tranform can be thought of as a 45° image rotation so that in addition to providing low rank structure, operating in (m-h) domain has the added benefit that the missing columns in each 354×354 matrix become missing diagonals in the transformed matrix. This allows our sampling scheme Ω the appropriate structure to succesfully achieve matrix completion (see [1] for details). Thus, in these experiments we require an additional modification of (1) by using the measurement operator $\mathcal{A}\mathcal{M}^*$ and modify (3) by utilizing $\mathcal{A}\mathcal{S}\mathcal{M}^*$. This way we achieve low rank structure in (m-h) domain, map our matrix back to our original (structured) domain and apply our sampling scheme to fit our unstructured oservations. These results are shown in Figure 5, where we plot the SNR of each reconstructed 354×354

frequency slice in terms of its frequency value. We then map our recovered data cube back into time domain, $\tilde{D}^\sharp \in \mathbb{C}^{1024\times 354\times 354}$, and fix a receiver location to observe a detailed subsection of the recovered 1024×354 common receiver gather in order to highlight the benefits of our methodology at the fine scale (also shown in Figure 5). For both $\rho = 1$ and $\rho = \frac{1}{2}$ the plots in Figure 5 show a clear advantage from the inclusion of the regularization operator, especially in the $7 - 65$ Hz frequency band where most of the energy of seismic wavefronts is found.

IV. Conclusions

We have presented and analyzed a low rank matrix recovery workflow that handles data corrupted by off-the-grid measurements. By utilizing the known deviated grid positions, we construct and incorporate a regularization operator that simulates the effect of the unstructured grid into the measurement model of our nuclear norm minimization problem. Analysis of our method and chosen regularization operator defined by trigonometric interpolation provide reconstruction error bounds that reflect the behaviour of the error depending on the degree of irregularity in our measurements.

Furthermore, we have presented matrix sensing and matrix completion numerical examples with unstructured observations that demonstrate the reconstruction SNR enhancements and an improved control on the approximation error when compared to the output of (1). We demonstrate numerically that these conclusions also extend to seismic trace interpolation, an application that illustrates the potential real world implications of our work.

Acknowledgements

The authors would like to thank Ernie Esser for his suggestions and stimulating discussions on the topics of this paper, it was a true priviledge to work with him. We miss him dearly.

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, DownUnder GeoSolutions, Hess Corporation, Petrobras, PGS, Sub Salt Solutions, WesternGeco, and Woodside. O. Ylmaz also acknowledges an NSERC Discovery Grant (22R82411) and an NSERC Accelerator Award (22R68054).

References

[1] Aleksandr Y. Aravkin, Rajiv Kumar, Hassan Mansour, Ben Recht and Felix J. Herrmann [2014], Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation. *SIAM Journal on Scientific Computing*, **36**, S237-S266.

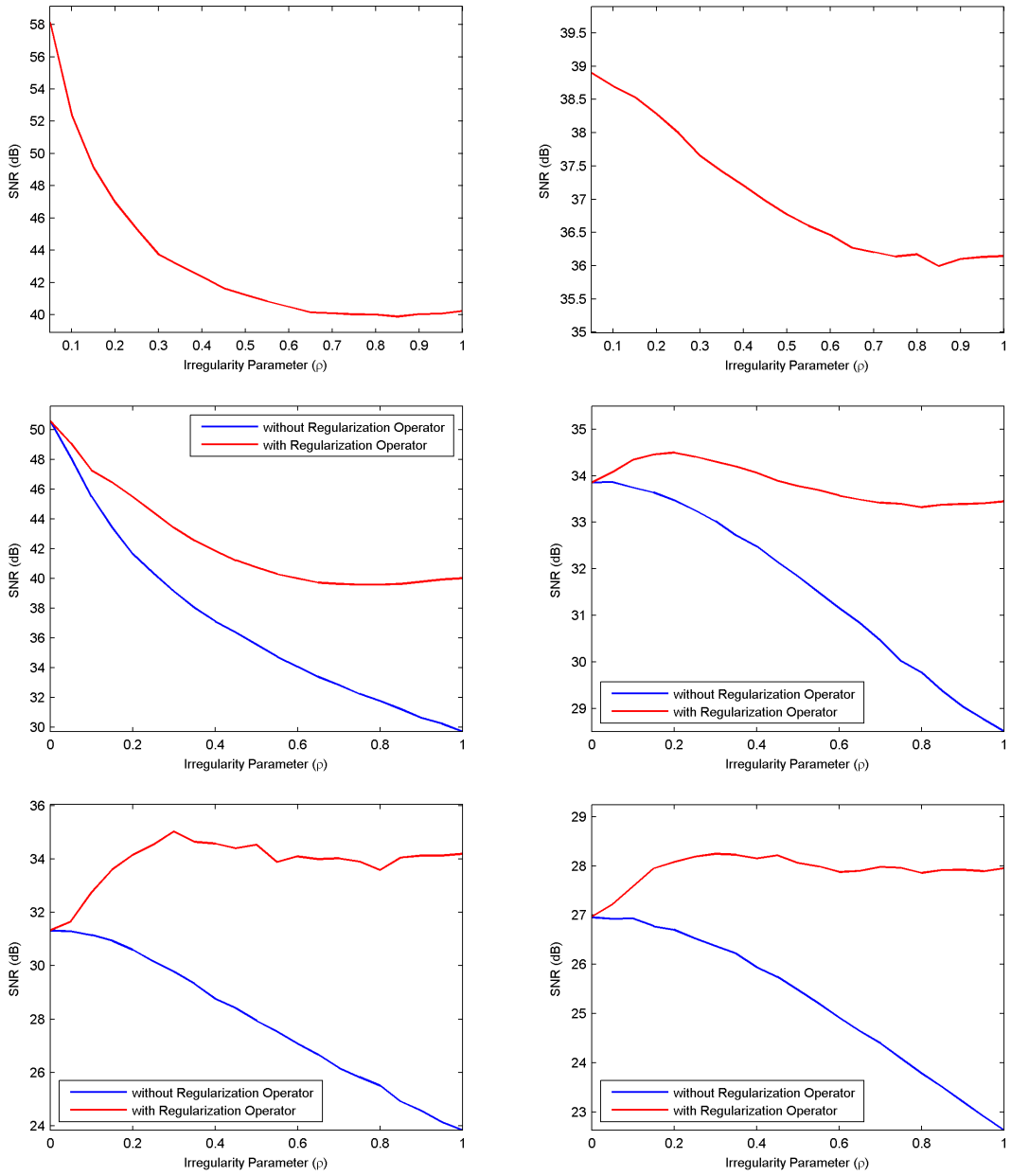

Fig. 4. Results of matrix completion experiments via LR-BPDN with the dataset introduced in Section III (see Figure 2). In each figure we plot the irregularity parameter ρ versus mean SNR of output of (1) or (3) or both and report the mean over 500 experiments, each with independent Ω realization. (top left) Data regularization experiments, i.e., $\Omega = [n] \times [m]$. (top right) Noisy data regularization experiments. (middle left) Matrix completion experiments with 50% missing entries. (middle right) Noisy matrix completion experiments with 50% missing entries. (bottom left) Matrix completion experiments with 75% missing entries. (middle right) Noisy matrix completion experiments with 75% missing entries.

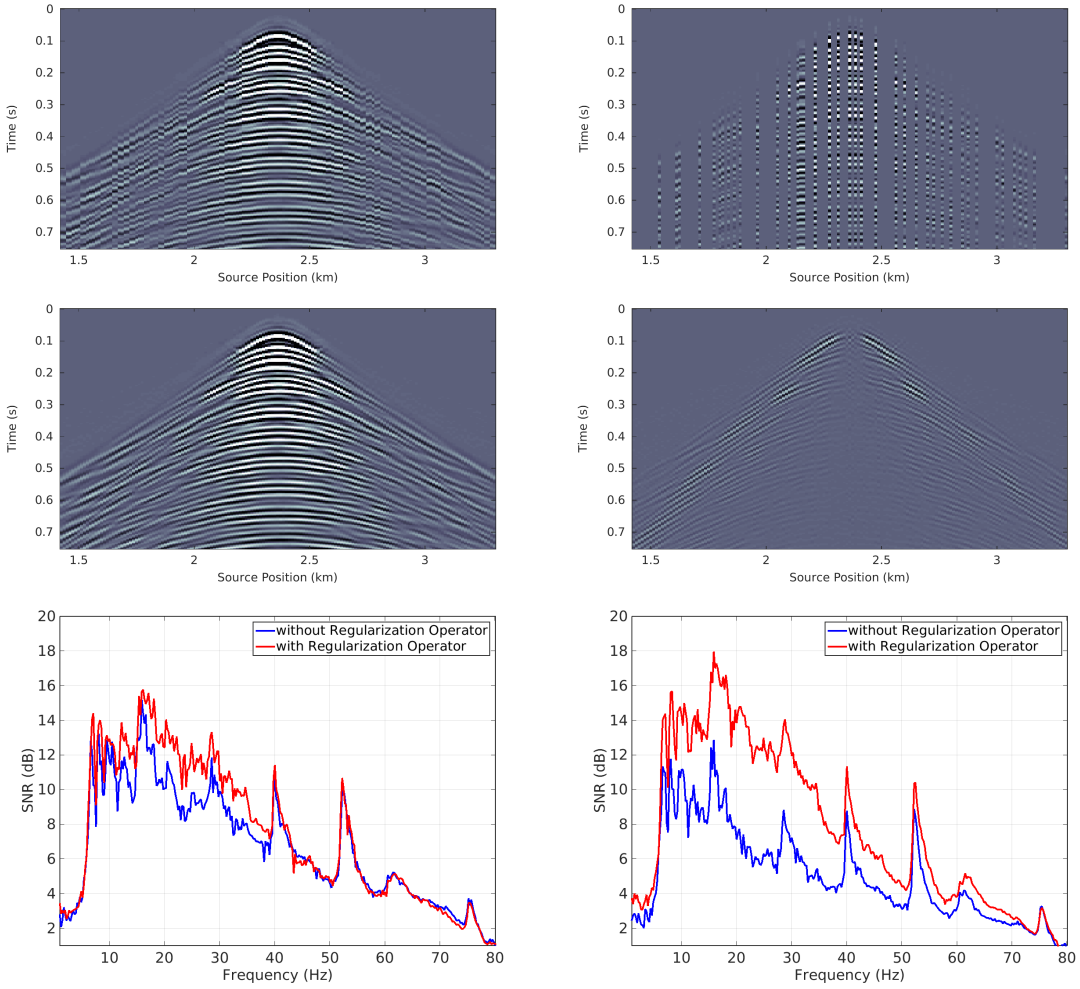

Fig. 5. (top) Subsection of unstructured (with $\rho = 1$) common receiver gather with receiver location at 2.4 km extracted from the irregularized Gulf of Suez dataset. (second row) Subsection of subsampled unstructured common receiver gather with 75% missing sources removed in a jittered random fashion. (third row) Subsection of interpolated and regularized common receiver gather via (3) with SNR = 14.9 dB. (fourth row) Subsection of difference plot of true structured common receiver gather and interpolated and regularized common receiver gather via (3). (fifth row) Plot of frequency (Hz) versus SNR of output of (1) and (3) matrix completion problem via LR-BPDN for Gulf of Suez dataset with irregularity parameter $\rho = \frac{1}{2}$. (bottom) Plot of frequency (Hz) versus SNR of output of (1) and (3) matrix completion problem via LR-BPDN for Gulf of Suez dataset with irregularity parameter $\rho = 1$.

[2] Benjamin Recht, Maryam Fazel, and Pablo A. Parillo [2010], Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review*, **52**, 471-501.

[3] Charles L. Epstein [2005], How well does the finite Fourier transform approximate the Fourier transform? *Communications on Pure and Applied Mathematics*, **58**, 14211435.

[4] Chengbo Li, Charles C. Mosher and Sam T. Kaplan [2012], Interpolated compressive sensing for seismic data reconstruction. *82nd annual international meeting, SEG, Expanded Abstracts*.

[5] Emmanuel J. Candès, and Benjamin Recht [2009], Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, **9**, 717-772.

[6] Emmanuel J. Candès, and Yaniv Plan [2009], Matrix Completion With Noise. *Proceedings of the IEEE* , **98**, 925 - 936.

[7] Ewout van den Berg and Michael P. Friedlander [2008], Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, **31(2)**, 890-912.

[8] Haneet Wason, Rajiv Kumar, Aleksandr Y. Aravkin and Felix J. Herrmann [2014], Source separation via SVD-free rank minimization in the hierarchical semi-separable representation. *SEG Technical Program Expanded Abstracts*, 120-126.

[9] Jens Keiner, Stefan Kunis and Daniel Potts [2008], Using NFFT 3 – a software library for various nonequispaced fast Fourier transforms. *ACM Trans. Math. Softw.*, **36**, 19:1–19:30.

[10] Jianwei Ma [2013]. Three-dimensional irregular seismic data reconstruction via low-rank matrix completion. *Geophysics*, **78** (5), V181-V192.

[11] Leslie Greengard and June-Yub Lee [2004], Accelerating the nonuniform fast Fourier transform. *Applied and Computational Harmonic Analysis*, **35**, 111-129.

[12] Maryam Fazel E. Candes, Benjamin Recht and Pablo A. Parrilo [2008], Compressed sensing and robust recovery of low rank matrices. *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, 1043-1047.

[13] Michael Pippig and Daniel Potts [2013], Parallel three-dimensional nonequispaced fast Fourier transforms and their applications to particle simulation. *SIAM Journal of Scientific Computing*, **35(4)**, C411-C437.

[14] Rajiv Kumar, Oscar López, Ernie Esser and Felix Herrmann [2015], Matrix completion on unstructured grids : 2-D seismic data regularization and interpolation. *EAGE Annual Conference Proceedings*.

[15] Simon Foucart and Holger Rauhut [2013], A Mathematical Introduction to Compressive Sensing. *Birkhäuser*.

[16] Theodore J. Rivlin [1969], An Introduction to the Approximation of Functions. *Blaisdell Publishing Company*.

[17] Tim T.Y. Lin and Felix Herrmann [2009], Designing simultaneous acquisitions with compressive sensing. *EAGE Annual Conference Proceedings*.

[18] Roman Vershynin [2012], Introduction to the non-asymptotic analysis of random matrices. *Cambridge University Press*.

[19] Yi Yang, Jianwei Ma and Stanley Osher [2012], Seismic data reconstruction via matrix completion. *UCLA CAM Report*, 12-14.

[20] Zhiyuan Weng and Xin Wang [2012], Low-rank matrix completion for array signal processing. *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2697-2700.

V. Proofs

We now provide proofs of all our Theorems and Lemmas.

A. Proof of Theorem II.1

For the proof of Theorem II.1 we first introduce some notation and lemmas.

For $\tilde{n}, \tilde{m} \in \mathbb{N}$ we define an exponential polynomial of order $(\tilde{n}, \tilde{m})$ as a function of the form

$$\mathbf{p}(x, y) = \sum_{k=-\tilde{n}}^{\tilde{n}} \sum_{\ell=-\tilde{m}}^{\tilde{m}} \alpha_{k,\ell} \mathbf{e}^{kx} \mathbf{e}^{\ell y},$$

for $\alpha_{k,\ell} \in \mathbb{C}$. Furthermore, let us denote the set of all such functions as $\mathcal{T}_{\tilde{n},\tilde{m}}$. For our function of interest, $\mathbf{D}$, we define $\mathbf{p}^* \in \mathcal{T}_{\tilde{n},\tilde{m}}$ as

$$\mathbf{p}^* = \underset{\mathbf{p} \in \mathcal{T}_{\tilde{n},\tilde{m}}}{\operatorname{argmin}} \|\mathbf{D} - \mathbf{p}\|_\infty,$$

so that $\mathbf{p}^*$ is the best uniform approximation to $\mathbf{D}$ from $\mathcal{T}_{\tilde{n},\tilde{m}}$. Notice that since $\mathcal{T}_{\tilde{n},\tilde{m}}$ is a finite-dimensional subspace of the normed linear space $C(Q)$, the existence of $\mathbf{p}^*$ is established.

We next consider the modulus of continutity of a function $\mathbf{f}$ in domain Q defined by

$$\omega_{\mathbf{f},Q}(\delta) := \sup_{\substack{x,y \in Q \\ \|x-y\|_2 \leq \delta}} |\mathbf{f}(x) - \mathbf{f}(y)|.$$

Notice that the modulus of continuity depends on the function and domain of interest, for brevity we will instead write $\omega(\delta)$ so that these terms are clear from the context.

Indeed this consideration makes sense for $\mathbf{D}$ in Q, since any Lipschitz continuous function admits a modulus of continuity. We will find it usefull to state properties of ω. First we can show that ω is increasing, so that if $0 < \delta_1 \leq \delta_2$ then

$$\omega(\delta_1) \leq \omega(\delta_2).$$

Furthermore, in a convex domain we have can establish subadditivity

$$\omega(\delta_1 + \delta_2) \leq \omega(\delta_1) + \omega(\delta_2)$$

for any $0 < \delta_1, \delta_2$. We can also show

$$\omega(\delta_1 \delta_2) \leq (1 + \delta_1)\omega(\delta_2).$$

Finally as an important observation notice that if $\mathbf{D}$ is Lipschitz continous in Q with Lipschitz constant L then we can write

$$\omega(\delta) \leq L\delta. \tag{8}$$

With these notations and observations, we may now introduce the following lemmas:

Lemma V.1. *Define $\mathcal{S}(D_0)(k, \ell)$ as in the Section II-A, then for any $k \in [n]$, $\ell \in [m]$ we can write*

$$\begin{aligned} &\mathcal{S}(D_0)(k, \ell) = \\ &\frac{1}{nm} \sum_{p=1}^{n} \sum_{q=1}^{m} \boldsymbol{D}(x^p, y^q) \boldsymbol{K}_{\tilde{n}}(\tilde{x}^{k,\ell} - x^p) \boldsymbol{K}_{\tilde{m}}(\tilde{y}^{k,\ell} - y^q). \end{aligned} \tag{9}$$

Where $\tilde{n} = \frac{n-1}{2}$ *and* $\tilde{m} = \frac{m-1}{2}$ *(since* n, m *were assumed to be odd) and for* $k \in \mathbb{N}$, $\boldsymbol{K}_k(\theta) = \frac{\sin(\pi(2k+1)\theta)}{\sin(\pi\theta)}$. $\{(x^p, y^q)\}_{p,q=1}^{n,m}$ *and* $\{(\tilde{x}^{k,\ell}, \tilde{y}^{k,\ell})\}_{k,\ell=1}^{n,m}$ *are our collection of equispaced and non-equispaced points respectively (see Section I-A).*

Lemma V.2. *Let* $\boldsymbol{p} \in \mathcal{T}_{\tilde{n},\tilde{m}}$*, then for any* $(x, y) \in Q$ *we can write*

$$\begin{aligned}&(2\tilde{n}+1)(2\tilde{m}+1)\boldsymbol{p}(x,y) = \\ &\sum_{p=1}^{2\tilde{n}+1}\sum_{q=1}^{2\tilde{m}+1} \boldsymbol{K}_{\tilde{n}}(x-x^p)\boldsymbol{K}_{\tilde{m}}(y-y^q)\boldsymbol{p}(x^p,y^q).\end{aligned} \tag{10}$$

Lemma V.3. *The best approximation* $\boldsymbol{p}^* \in \mathcal{T}_{\tilde{n},\tilde{m}}$ *of* $\boldsymbol{D}$ *in* Q *satisfies*

$$\|\boldsymbol{D} - \boldsymbol{p}^*\|_\infty \leq 6\omega\left(\frac{1}{2\pi\tilde{n}}\right) + 6\omega\left(\frac{1}{2\pi\tilde{m}}\right), \tag{11}$$

where the modulus of continuity is considered with respect to $\boldsymbol{D}$ *in* Q*.*

We now proceed to the proof of Theorem II.1.

Proof of Theorem II.1. We begin by applying the triangle inequality

$$\begin{aligned}&|\mathcal{S}(D_0)(k,\ell) - D_\rho(k,\ell)| = |\mathcal{S}(D_0)(k,\ell) - \mathbf{D}(\tilde{x}^{k,\ell}, \tilde{y}^{k,\ell})| \\ &\leq |\mathcal{S}(D_0)(k,\ell) - \mathbf{p}^*(\tilde{x}^{k,\ell}, \tilde{y}^{k,\ell})| + \|\mathbf{p}^* - \mathbf{D}\|_\infty.\end{aligned}$$

Focusing on the term $|\mathcal{S}(D_0)(k,\ell) - \mathbf{p}^*(\tilde{x}^{k,\ell}, \tilde{y}^{k,\ell})|$, since $\mathbf{p}^* \in \mathcal{T}_{\tilde{n},\tilde{m}}$ we apply (9) and (10) with $\tilde{n} = \frac{n-1}{2}$, $\tilde{m} = \frac{m-1}{2}$ (since n and m were assumed to be odd in Section II-A) and the triangle inequality to obtain

$$\begin{aligned}&|\mathcal{S}(D_0)(k,\ell) - \mathbf{p}^*(\tilde{x}^{k,\ell}, \tilde{y}^{k,\ell})| \leq \\ &\sum_{p=1}^{2\tilde{n}+1}\sum_{q=1}^{2\tilde{m}+1} \frac{\|\mathbf{D} - \mathbf{p}^*\|_\infty |\mathbf{K}_{\tilde{n}}(\tilde{x}^{k,\ell} - x^p)\mathbf{K}_{\tilde{m}}(\tilde{y}^{k,\ell} - y^q)|}{(2\tilde{n}+1)(2\tilde{m}+1)}.\end{aligned} \tag{12}$$

We now consider the summation:

$$S_{\tilde{n}} := \sum_{p=1}^{2\tilde{n}+1} |\mathbf{K}_{\tilde{n}}(\tilde{x}^{k,\ell} - x^p)|$$

By definition of $\mathbf{K}_{\tilde{n}}$ and since $\tilde{x}^{k,\ell} - x^p = \frac{k-p}{2\tilde{n}+1} + \Delta(k,\ell)$, we obtain

$$\begin{aligned} S_{\tilde{n}} &= \sum_{p=1}^{2\tilde{n}+1} \frac{|\sin(\pi(k-p) + \pi(2\tilde{n}+1)\Delta(k,\ell))|}{|\sin(\pi\frac{k-p}{2\tilde{n}+1} + \pi\Delta(k,\ell))|} \\ &= \sum_{p=1}^{2\tilde{n}+1} \frac{|\sin(\pi(2\tilde{n}+1)\Delta(k,\ell))|}{|\sin(\pi\frac{k-p}{2\tilde{n}+1} + \pi\Delta(k,\ell))|} \\ &= \sum_{p=0}^{2\tilde{n}} \frac{|\sin(\pi(2\tilde{n}+1)\Delta(k,\ell))|}{|\sin(\pi\frac{p}{2\tilde{n}+1} + \pi\Delta(k,\ell))|}. \end{aligned}$$

Where the second equality follows from angle sum indentities and the last equality follows by using the fact that $\frac{1}{|\sin(\pi x)|}$ is 1-periodic. We now bound the summation $\frac{1}{(2\tilde{n}+1)}\sum_{p=0}^{2\tilde{n}} \frac{1}{|\sin(\pi\frac{p}{2\tilde{n}+1}+\pi\Delta(k,\ell))|}$ by considering the four cases: $\frac{1}{2}\frac{1}{2\tilde{n}+1} \leq \Delta(k,\ell) < \frac{1}{2\tilde{n}+1}$, $0 < \Delta(k,\ell) < \frac{1}{2}\frac{1}{2\tilde{n}+1}$, $-\frac{1}{2}\frac{1}{2\tilde{n}+1} < \Delta(k,\ell) < 0$ and $-\frac{1}{2\tilde{n}+1} < \Delta(k,\ell) \leq -\frac{1}{2}\frac{1}{2\tilde{n}+1}$. Let us deal with the following case:

Assume $\frac{1}{2}\frac{1}{2\tilde{n}+1} \leq \Delta(k,\ell) < \frac{1}{2\tilde{n}+1}$. Since $\frac{1}{|\sin(\pi x)|}$ is decreasing for $x \in [0, \frac{\pi}{2}]$ and increasing for $x \in [\frac{\pi}{2}, \pi]$, we separate our sum over the indices accordingly and bound each summation with an appropriate integral. Concluding we obtain

$$\begin{aligned} &\sum_{p=0}^{2\tilde{n}} \frac{1}{|\sin(\pi\frac{p}{2\tilde{n}+1} + \pi\Delta(k,\ell))|} \\ \leq &\frac{1}{|\sin(\pi\Delta(k,\ell))|} + \frac{1}{|\sin(\pi\frac{2\tilde{n}}{2\tilde{n}+1} + \pi\Delta(k,\ell))|} \\ &+(2\tilde{n}+1)\int_0^{\frac{2\tilde{n}}{2\tilde{n}+1}} \frac{dx}{|\sin(\pi(x+\Delta(k,\ell)))|} \\ = &\frac{1}{|\sin(\pi\Delta(k,\ell))|} + \frac{1}{|\sin(\pi\frac{2\tilde{n}}{2\tilde{n}+1} + \pi\Delta(k,\ell))|} \\ &+(2\tilde{n}+1)\frac{1}{\pi}\log\left(\frac{\tan(\frac{\pi}{2}(\frac{2\tilde{n}}{2\tilde{n}+1} + |\Delta(k,\ell)|))}{\tan(\frac{\pi}{2}|\Delta(k,\ell)|)}\right). \end{aligned}$$

By observing that $\frac{|\sin(\pi(2\tilde{n}+1)\Delta(k,\ell))|}{(2\tilde{n}+1)|\sin(\pi\Delta(k,\ell))|} \leq 1$ and $\frac{|\sin(\pi(2\tilde{n}+1)\Delta(k,\ell))|}{(2\tilde{n}+1)|\sin(\pi\frac{2\tilde{n}}{2\tilde{n}+1}+\pi\Delta(k,\ell))|} \leq 1$, we have shown

$$\frac{S_{\tilde{n}}}{2\tilde{n}+1} \leq \mathbf{g}_{2\tilde{n}+1}(\Delta(k,\ell)) \tag{13}$$

where $\mathbf{g}_{2\tilde{n}+1}$ is as defined in Section II-A. For the remaining allowed values of $\Delta(k,\ell)$ we can use similar arguments to obtain the same conclusion. Furthermore, we may also adopt this line of reasoning to obtain

$$\frac{1}{(2\tilde{m}+1)}\sum_{q=1}^{2\tilde{m}+1} |\mathbf{K}_{\tilde{m}}(\tilde{y}^{k,\ell} - y^q)| \leq \mathbf{g}_{2\tilde{m}+1}(\Gamma(k,\ell)).$$

Combining all of this with (16) and the beginning of the proof, we see that:

$$|\mathcal{S}(D_0)(k,\ell) - D_\rho(k,\ell)|$$
$$\leq \|\mathbf{p}^* - \mathbf{D}\|_\infty [1 + \mathbf{g}_n(\Delta(k,\ell))\mathbf{g}_m(\Gamma(k,\ell))].$$

Finally applying (11) and our observation (8), we conclude:

$$|\mathcal{S}(D_0)(k,\ell) - D_\rho(k,\ell)| \leq$$
$$c_0 + c_0 \mathbf{g}_n(\Delta(k,\ell))\mathbf{g}_m(\Gamma(k,\ell)).$$

where $c_0 = \left(\frac{6L}{\pi(n-1)} + \frac{6L}{\pi(m-1)}\right)$.

□

B. Proof of Lemmas V.1, V.2 and V.3

Proof of Lemma V.1. We begin by stating the well known calculation

$$\sum_{k=-\ell}^{\ell} \mathbf{e}^{k\theta} = \frac{\sin(\pi(2\ell+1)\theta)}{\sin(\pi\theta)} := \mathbf{K}_\ell(\theta), \tag{14}$$

for any $\ell \in \mathbb{N}$, which follows from properties of geometric series.

Now we use our observations in Section II-A with $\tilde{n} = \frac{n-1}{2}$ and $\tilde{m} = \frac{m-1}{2}$ (since n and m were assumed to be odd) to write for any $k \in [n]$ and $\ell \in [m]$

$$\begin{aligned}
&(nm)\mathcal{S}(D_0)(k,\ell) \\
=&(nm)\sum_{u=1}^{n}\sum_{v=1}^{m} \check{D}_0(u,v)\mathbf{e}^{-(u-\tilde{n}-1)\tilde{x}^{k,\ell}}\mathbf{e}^{-(v-\tilde{m}-1)\tilde{y}^{k,\ell}} \\
=&\sum_{u=-\tilde{n}}^{\tilde{n}}\sum_{v=-\tilde{m}}^{\tilde{m}}\sum_{p=1}^{n}\sum_{q=1}^{m} D_0(p,q)\mathbf{e}^{ux^p}\mathbf{e}^{vy^q}\mathbf{e}^{-u\tilde{x}^{k,\ell}}\mathbf{e}^{-v\tilde{y}^{k,\ell}} \\
=&\sum_{p=1}^{n}\sum_{q=1}^{m} D_0(p,q)\sum_{u=-\tilde{n}}^{\tilde{n}}\sum_{v=-\tilde{m}}^{\tilde{m}} \mathbf{e}^{u(x^p-\tilde{x}^{k,\ell})}\mathbf{e}^{v(y^q-\tilde{y}^{k,\ell})} \\
=&\sum_{p=1}^{n}\sum_{q=1}^{m} \mathbf{D}(x^p,y^q)\mathbf{K}_{\tilde{n}}(\tilde{x}^{k,\ell}-x^p)\mathbf{K}_{\tilde{m}}(\tilde{y}^{k,\ell}-y^q).
\end{aligned}$$

Where second equality holds by using the definition of $\check{D}_0$ in Section II-A and changing the summations over u and v to run from $-\tilde{n}$ to $\tilde{n}$ and $-\tilde{m}$ to $\tilde{m}$ respectively. The last equality holds from (14) and symmetry over the indices. □

Proof of Lemma V.2. Since $\mathbf{p} \in \mathcal{T}_{\tilde{n},\tilde{m}}$, let us write

$$\mathbf{p}(x,y) = \sum_{k=-\tilde{n}}^{\tilde{n}} \sum_{\ell=-\tilde{m}}^{\tilde{m}} \alpha_{k,\ell} \mathbf{e}^{kx} \mathbf{e}^{\ell y},$$

for some $\alpha_{k,\ell} \in \mathbb{C}$. Focusing on the summation terms on the right hand side of (10), let us fix $p \in [2\tilde{n}+1]$ and apply (14) to obtain

$$\begin{aligned}
&\sum_{q=1}^{2\tilde{m}+1} \mathbf{K}_{\tilde{m}}(y-y^q)\mathbf{p}(x^p,y^q) \\
&= \sum_{q=1}^{2\tilde{m}+1} \left(\sum_{\tilde{\ell}=-\tilde{m}}^{\tilde{m}} \mathbf{e}^{\tilde{\ell}(y-y^q)} \right) \left(\sum_{k=-\tilde{n}}^{\tilde{n}} \sum_{\ell=-\tilde{m}}^{\tilde{m}} \alpha_{k,\ell} \mathbf{e}^{kx^p} \mathbf{e}^{\ell y^q} \right) \\
&= \sum_{q=1}^{2\tilde{m}+1} \sum_{k=-\tilde{n}}^{\tilde{n}} \mathbf{e}^{kx^p} \left(\sum_{\ell=-\tilde{m}}^{\tilde{m}} \alpha_{k,\ell} \mathbf{e}^{\ell y^q} \right) \left(\sum_{\tilde{\ell}=-\tilde{m}}^{\tilde{m}} \mathbf{e}^{\tilde{\ell}(y-y^q)} \right) \\
&= \sum_{q=1}^{2\tilde{m}+1} \sum_{k=-\tilde{n}}^{\tilde{n}} \mathbf{e}^{kx^p} \sum_{\ell=-\tilde{m}}^{\tilde{m}} \alpha_{k,\ell} \left(\mathbf{e}^{\ell y} + \sum_{\tilde{\ell} \neq \ell} \mathbf{e}^{y^q(\ell-\tilde{\ell})} \mathbf{e}^{\tilde{\ell} y} \right) \\
&= \sum_{k=-\tilde{n}}^{\tilde{n}} \sum_{\ell=-\tilde{m}}^{\tilde{m}} \mathbf{e}^{kx^p} \alpha_{k,\ell} \sum_{q=1}^{2\tilde{m}+1} \left(\mathbf{e}^{\ell y} + \sum_{\tilde{\ell} \neq \ell} \mathbf{e}^{y^q(\ell-\tilde{\ell})} \mathbf{e}^{\tilde{\ell} y} \right).
\end{aligned}$$

We now distribute the sum over q, use the definition of y^q and the orthogonality of the exponential basis functions to see that $\sum_{q=1}^{2\tilde{m}+1} \mathbf{e}^{y^q(\ell-\tilde{\ell})} = 0$ for any $-\tilde{n} \leq \ell, \tilde{\ell} \leq \tilde{n}$ with $\ell \neq \tilde{\ell}$, so that

$$\begin{aligned}
&\sum_{q=1}^{2\tilde{m}+1} \mathbf{K}_{\tilde{m}}(y-y^q)\mathbf{p}(x^p,y^q) \\
&= \sum_{k=-\tilde{n}}^{\tilde{n}} \sum_{\ell=-\tilde{m}}^{\tilde{m}} \mathbf{e}^{kx^p} \alpha_{k,\ell} (2\tilde{m}+1) \mathbf{e}^{\ell y}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\frac{1}{2\tilde{m}+1} \sum_{p=1}^{2\tilde{n}+1} \sum_{q=1}^{2\tilde{n}+1} \mathbf{K}_{\tilde{n}}(x-x^p)\mathbf{K}_{\tilde{m}}(y-y^q)\mathbf{p}(x^p,y^q) \\
&= \sum_{p=1}^{2\tilde{n}+1} \left(\sum_{\tilde{k}=-\tilde{n}}^{\tilde{n}} \mathbf{e}^{\tilde{k}(x-x^p)} \right) \sum_{k=-\tilde{n}}^{\tilde{n}} \sum_{\ell=-\tilde{m}}^{\tilde{m}} \mathbf{e}^{kx^p} \alpha_{k,\ell} \mathbf{e}^{\ell y}.
\end{aligned}$$

We may now rearrange the terms and use the same argument over indices $k, \tilde{k}$ and p to conclude

$$\begin{aligned}
&\sum_{p=1}^{2\tilde{n}+1} \sum_{q=1}^{2\tilde{n}+1} \frac{\mathbf{K}_{\tilde{n}}(x-x^p)\mathbf{K}_{\tilde{m}}(y-y^q)\mathbf{p}(x^p,y^q)}{(2\tilde{n}+1)(2\tilde{m}+1)} \\
&= \sum_{k=1}^{2\tilde{n}+1} \sum_{\ell=1}^{2\tilde{n}+1} \alpha_{k,\ell} \mathbf{e}^{kx} \mathbf{e}^{\ell y} = \mathbf{p}(x,y),
\end{aligned}$$

as desired.

□

Proof of Lemma V.3. For this proof we utilize analysis similar to Theorem 1.3 in [16], generalizing to continuous functions in 2 dimensions. For the time being we consider our function of interest $\mathbf{D}(x,y)$: $\tilde{Q} := [-\pi,\pi]^2 \mapsto \mathbb{C}$ with compact support and let $\tilde{\mathbf{D}}$ be its periodic extension to the whole plane. Next we define

$$\begin{aligned}\mathbf{S}_{\tilde{n},\tilde{m}}(x,y) := \sum_{k=0}^{\tilde{n}}\sum_{\ell=0}^{\tilde{m}} \rho_k\tau_\ell(&\alpha_{k,\ell}\cos(kx)\cos(\ell y)\\ &+\beta_{k,\ell}\cos(kx)\sin(\ell y)\\ &+\gamma_{k,\ell}\sin(kx)\cos(\ell y)\\ &+\kappa_{k,\ell}\sin(kx)\sin(\ell y))\end{aligned}$$

where $\rho_0 = \tau_0 = \frac{1}{2}$, ρ_k, τ_ℓ are real numbers for $k,\ell \geq 1$ to be determined and $\alpha_{k,\ell}, \beta_{k,\ell}, \gamma_{k,\ell}, \kappa_{k,\ell} \in \mathbb{C}$ are the respective Fourier coefficients of $\tilde{\mathbf{D}}$. In particular notice that $\mathbf{S}_{\tilde{n},\tilde{m}} \in \mathcal{T}_{\tilde{n},\tilde{m}}$, so that our aim is to bound $\left|\tilde{\mathbf{D}}(x,y) - \mathbf{S}_{\tilde{n},\tilde{m}}(x,y)\right|$ for any $x,y \in [-\pi,\pi]$. This bound in turn will apply to the best uniform approximation $\mathbf{p}^* \in \mathcal{T}_{\tilde{n},\tilde{m}}$.

Substituting the definition of the Fourier coefficients, $\alpha_{k,\ell}, \beta_{k,\ell}, \gamma_{k,\ell}, \kappa_{k,\ell}$, and using the trigonometric identity $\cos(A)\cos(B) + \sin(A)\sin(B) = \cos(A-B)$ and the fact that $\tilde{\mathbf{D}}$ is periodic, we can show that

$$\begin{aligned}&\mathbf{S}_{\tilde{n},\tilde{m}}(x,y) =\\ &\frac{1}{\pi^2}\iint_{\tilde{Q}} \tilde{\mathbf{D}}(\theta+x,\eta+y)\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)\,d\theta\,d\eta\end{aligned}$$

where

$$\begin{aligned}&\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta) :=\\ &\left(\rho_0 + \sum_{k=1}^{\tilde{n}}\rho_k\cos(k\theta)\right)\left(\tau_0 + \sum_{\ell=1}^{\tilde{m}}\tau_\ell\cos(\ell\eta)\right).\end{aligned} \tag{15}$$

For now let us assume that we can choose ρ_k, τ_ℓ for $k,\ell \geq 1$ so that $\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta) \geq 0$ in $[-\pi,\pi]^2$, and notice that

$$I_0 := \frac{1}{\pi^2}\iint_{\tilde{Q}} \mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)d\theta d\eta = 4\rho_0\tau_0 = 1.$$

Our observations so far give us

$$\begin{aligned}&\left|\tilde{\mathbf{D}}(x,y)-\mathbf{S}_{\tilde{n},\tilde{m}}(x,y)\right|\leq\\&\frac{1}{\pi^2}\iint_{\tilde{Q}}\left|\tilde{\mathbf{D}}(x,y)-\tilde{\mathbf{D}}(\theta+x,\eta+y)\right|\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)\,d\theta\,d\eta.\end{aligned}$$

We now utilize the modulus of continuity of $\tilde{\mathbf{D}}$ in $[-\pi,\pi]^2$. From our stated properties of the modulus of continuity above, we see that

$$\begin{aligned}&\left|\tilde{\mathbf{D}}(x,y)-\tilde{\mathbf{D}}(\theta+x,\eta+y)\right|\leq\omega(\|(\theta,\eta)\|_2)\\&\leq\omega(|\theta|)+\omega(|\eta|)=\omega\left(\tilde{n}|\theta|\frac{1}{\tilde{n}}\right)+\omega\left(\tilde{m}|\eta|\frac{1}{\tilde{m}}\right)\\&\leq(1+\tilde{n}|\theta|)\omega\left(\frac{1}{\tilde{n}}\right)+(1+\tilde{m}|\eta|)\omega\left(\frac{1}{\tilde{m}}\right).\end{aligned}$$

Combining our results above, we have

$$\begin{aligned}&|\tilde{\mathbf{D}}(x,y)-\mathbf{S}_{\tilde{n},\tilde{m}}(x,y)|\\&\leq\omega\left(\frac{1}{\tilde{n}}\right)+\omega\left(\frac{1}{\tilde{n}}\right)\frac{\tilde{n}}{\pi^2}\iint_{\tilde{Q}}|\theta|\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)\,d\theta\,d\eta\\&+\omega\left(\frac{1}{\tilde{m}}\right)+\omega\left(\frac{1}{\tilde{m}}\right)\frac{\tilde{m}}{\pi^2}\iint_{\tilde{Q}}|\eta|\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)\,d\theta\,d\eta.\end{aligned}$$

Let us focus on the last integral above and use Schwarz inequality to obtain

$$\begin{aligned}&\frac{1}{\pi^2}\iint_{\tilde{Q}}|\eta|\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)\,d\theta\,d\eta\\=&\frac{1}{\pi^2}\iint_{\tilde{Q}}(|\eta|\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)^{\frac{1}{2}})\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)^{\frac{1}{2}}\,d\theta\,d\eta\\\leq&\left(\frac{1}{\pi^2}\iint_{\tilde{Q}}|\eta|^2\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)\,d\theta\,d\eta\right)^{\frac{1}{2}}\sqrt{I_0}.\end{aligned}$$

Next, since $|\eta|\leq\frac{\pi}{2}\sin(|\eta|)$ for $|\eta|\leq\frac{\pi}{2}$ and $\frac{|\eta|}{2}\leq\frac{\pi}{2}$ in $[-\pi,\pi]$, we see that

$$\left(\frac{|\eta|}{2}\right)^2\leq\frac{\pi^2}{4}\sin^2\left(\frac{|\eta|}{2}\right)=\frac{\pi^2}{8}(1-\cos(|\eta|)).$$

Thus

$$\begin{aligned}&\left[\frac{1}{\pi^2}\iint_{\tilde{Q}}|\eta|^2\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)\,d\theta\,d\eta\right]^{\frac{1}{2}}\\&\leq\left[\frac{1}{2}\iint_{\tilde{Q}}(1-\cos(|\eta|))\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)\,d\theta\,d\eta\right]^{\frac{1}{2}}\\&=\left[\frac{\pi^2}{2}\sqrt{I_0}-\frac{1}{2}\iint_{\tilde{Q}}\rho_0\tau_1\cos^2(\eta)\,d\theta\,d\eta\right]^{\frac{1}{2}}\\&=\frac{\pi}{\sqrt{2}}\sqrt{1-\tau_1}.\end{aligned}$$

Where the second-to-last equality holds by distributing $-\cos(|\eta|)$ into the expression of $\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)$ and using orthogonality relations. Similarly we can show

$$\frac{1}{\pi^2}\iint_{\tilde{Q}} |\theta|\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta)\,d\theta\,d\eta \leq \frac{\pi}{\sqrt{2}}\sqrt{1-\rho_1},$$

so that we have concluded

$$\begin{aligned}
&\left|\tilde{\mathbf{D}}(x,y) - \mathbf{S}_{\tilde{n},\tilde{m}}(x,y)\right| \\
&\leq \omega\left(\frac{1}{\tilde{n}}\right)\left(1+\tilde{n}\frac{\pi}{\sqrt{2}}\sqrt{1-\rho_1}\right) \\
&+\omega\left(\frac{1}{\tilde{m}}\right)\left(1+\tilde{m}\frac{\pi}{\sqrt{2}}\sqrt{1-\tau_1}\right).
\end{aligned} \tag{16}$$

We will finish once we can choose ρ_k, τ_ℓ for $k,\ell \geq 1$ so that $\mathbf{U}_{\tilde{n},\tilde{m}}(\theta,\eta) \geq 0$ in $[-\pi,\pi]^2$, which in turn will determine ρ_1 and τ_1 above and finish the proof. For this task we produce our non-negative cosine polynomial similar to [16] in the discussion of Jackson's theorem. To this end, notice that for any sequences of real numbers $\{c_k\}_{k=1}^{\tilde{n}}, \{b_\ell\}_{\ell=1}^{\tilde{m}}$ and $\theta,\eta \in [-\pi,\pi]$ we have

$$0 \leq \left|\sum_{k=0}^{\tilde{n}} c_k e^{\mathbb{i}k\theta}\right|^2 \left|\sum_{\ell=0}^{\tilde{m}} b_\ell e^{\mathbb{i}\ell\eta}\right|^2 =$$
$$\left(\sum_{k=0}^{\tilde{n}} c_k e^{\mathbb{i}k\theta}\sum_{\tilde{k}=0}^{\tilde{n}} c_{\tilde{k}} e^{-\mathbb{i}\tilde{k}\theta}\right)\left(\sum_{\ell=0}^{\tilde{m}} b_\ell e^{\mathbb{i}\ell\eta}\sum_{\tilde{\ell}=0}^{\tilde{m}} b_{\tilde{\ell}} e^{-\mathbb{i}\tilde{\ell}\eta}\right)$$

and furthermore

$$\begin{aligned}
C &:= \left(\sum_{k=0}^{\tilde{n}} c_k e^{\mathbb{i}k\theta}\right)\left(\sum_{\tilde{k}=0}^{\tilde{n}} c_{\tilde{k}} e^{-\mathbb{i}\tilde{k}\theta}\right) \\
&= \sum_{\tilde{k}=0}^{\tilde{n}} c_{\tilde{k}}^2 + \sum_{\tilde{k}=0}^{\tilde{n}-1} c_{\tilde{k}}c_{\tilde{k}+1}\left(e^{\mathbb{i}\theta}+e^{-\mathbb{i}\theta}\right) \\
&+ \sum_{\tilde{k}=0}^{\tilde{n}-2} c_{\tilde{k}}c_{\tilde{k}+2}\left(e^{\mathbb{i}2\theta}+e^{-\mathbb{i}2\theta}\right) \\
&+ \cdots + \sum_{\tilde{k}=0}^{\tilde{n}-k} c_{\tilde{k}}c_{\tilde{k}+k}\left(e^{\mathbb{i}k\theta}+e^{-\mathbb{i}k\theta}\right) \\
&+ \cdots + c_0 c_{\tilde{n}}\left(e^{\mathbb{i}\tilde{n}\theta}+e^{-\mathbb{i}\tilde{n}\theta}\right) \\
&= \sum_{\tilde{k}=0}^{\tilde{n}} c_{\tilde{k}}^2 + \sum_{k=1}^{\tilde{n}}\sum_{\tilde{k}=0}^{\tilde{n}-k} c_{\tilde{k}}c_{\tilde{k}+k}\left(e^{\mathbb{i}k\theta}+e^{-\mathbb{i}k\theta}\right) \\
&= \sum_{\tilde{k}=0}^{\tilde{n}} c_{\tilde{k}}^2 + \sum_{k=1}^{\tilde{n}}\sum_{\tilde{k}=0}^{\tilde{n}-k} c_{\tilde{k}}c_{\tilde{k}+k} 2\cos(k\theta).
\end{aligned}$$

We can obtain similar results for the sums over $\{b_\ell\}_{\ell=1}^{\tilde{m}}$, so that

$$0 \leq C \left(\sum_{\tilde{\ell}=0}^{\tilde{m}} b_{\tilde{\ell}}^2 + \sum_{\ell=1}^{\tilde{m}} \sum_{\tilde{\ell}=0}^{\tilde{m}-\ell} b_{\tilde{\ell}} b_{\tilde{\ell}+\ell} 2\cos(\ell\eta) \right).$$

By definition of $\mathbf{U}_{\tilde{n},\tilde{m}}$ (15), we accomplish our goal if we define

$$\rho_0 = \sum_{\tilde{k}=0}^{\tilde{n}} c_{\tilde{k}}^2 \text{ and } \tau_0 = \sum_{\tilde{\ell}=0}^{\tilde{m}} b_{\tilde{\ell}}^2$$

and for $\tilde{k}, \tilde{\ell} \geq 1$

$$\rho_k = 2\sum_{\tilde{k}=0}^{\tilde{n}-k} c_{\tilde{k}} c_{\tilde{k}+k} \text{ and } \tau_\ell = 2\sum_{\tilde{\ell}=0}^{\tilde{m}-\ell} b_{\tilde{\ell}} b_{\tilde{\ell}+\ell}.$$

However, we have chosen constraints $\rho_0 = \tau_0 = \frac{1}{2}$. Since the bound (16) is tighter if $\rho_1, \tau_1 \approx 1$, let us choose $c_k = c\sin(\frac{k+1}{\tilde{n}+2}\pi)$ and $b_\ell = b\sin(\frac{\ell+1}{\tilde{m}+2}\pi)$ where c and b are our normalizing constants $c^2 = \frac{1}{2}(\sum_{k=0}^{\tilde{n}} \sin^2(\frac{k+1}{\tilde{n}+2}\pi))^{-1}$ and $b^2 = \frac{1}{2}(\sum_{\ell=0}^{\tilde{m}} \sin^2(\frac{\ell+1}{\tilde{m}+2}\pi))^{-1}$. This choice trivially satisfies our constraints. Notice that

$$\rho_1 = 2c^2 \sum_{k=0}^{\tilde{n}} \sin\left(\frac{k+1}{\tilde{n}+2}\pi\right) \sin\left(\frac{k+2}{\tilde{n}+2}\pi\right) := S_1,$$

$$\rho_1 = 2c^2 \sum_{k=0}^{\tilde{n}} \sin\left(\frac{k}{\tilde{n}+2}\pi\right) \sin\left(\frac{k+1}{\tilde{n}+2}\pi\right) := S_2,$$

so that

$$\rho_1 = S_1 + S_2 \tag{17}$$

$$= 2c^2 \cos\left(\frac{\pi}{\tilde{n}+2}\right) \sum_{k=0}^{\tilde{n}} \sin^2\left(\frac{k+1}{\tilde{n}+2}\pi\right) \tag{18}$$

$$= 2c^2 \cos\left(\frac{\pi}{\tilde{n}+2}\right) \frac{1}{2c^2} = \cos\left(\frac{\pi}{\tilde{n}+2}\right). \tag{19}$$

The second equality holds by applying the identity $\sin(A) + \sin(B) = 2\sin\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right)$. We can similarly show that $\tau_1 = \cos\left(\frac{\pi}{\tilde{m}+2}\right)$, so that

$$\begin{aligned}
&\left|\tilde{\mathbf{D}}(x,y) - \mathbf{S}_{\tilde{n},\tilde{m}}(x,y)\right| \\
&\leq \omega\left(\frac{1}{\tilde{n}}\right)\left(1 + \tilde{n}\frac{\pi}{\sqrt{2}}\sqrt{1-\cos\left(\frac{\pi}{\tilde{n}+2}\right)}\right) \\
&+ \omega\left(\frac{1}{\tilde{m}}\right)\left(1 + \tilde{m}\frac{\pi}{\sqrt{2}}\sqrt{1-\cos\left(\frac{\pi}{\tilde{m}+2}\right)}\right).
\end{aligned}$$

Finally, we note that

$$\tilde{n}\sqrt{1-\cos\left(\frac{\pi}{\tilde{n}+2}\right)} = \tilde{n}\sqrt{2}\sin\left(\frac{\pi}{2(\tilde{n}+2)}\right)$$
$$\leq \tilde{n}\sqrt{2}\frac{\pi}{2(\tilde{n}+2)} \leq \frac{\pi}{\sqrt{2}} \leq \sqrt{5}$$

and similarly $\tilde{m}\sqrt{1-\cos\left(\frac{\pi}{\tilde{m}+2}\right)} \leq \sqrt{5}$. An additional application of $\frac{\pi}{\sqrt{2}} \leq \sqrt{5}$ gives

$$\left|\tilde{\mathbf{D}}(x,y) - \mathbf{S}_{\tilde{n},\tilde{m}}(x,y)\right| \leq 6\omega\left(\frac{1}{\tilde{n}}\right) + 6\omega\left(\frac{1}{\tilde{m}}\right).$$

This gives the results for $\mathbf{D}(x,y) : \tilde{Q} \mapsto \mathbb{C}$. By applying a linear transformation, we can map $\tilde{Q} := [-\pi,\pi]^2$ to our domain of interest $Q = [-\frac{1}{2},\frac{1}{2}]^2$ and show that for $\mathbf{D}(x,y) : Q \mapsto \mathbb{C}$

$$|\mathbf{D}(x,y) - \mathbf{S}_{\tilde{n},\tilde{m}}(x,y)| \leq 6\omega\left(\frac{1}{2\pi\tilde{n}}\right) + 6\omega\left(\frac{1}{2\pi\tilde{m}}\right)$$

holds, where the modulus of continuity is taken with respect to $\mathbf{D}$ in Q. □

C. Proof of Theorem I.1

Proof of Theorem I.1. If we recall our notations from Section II, we begin by noticing that for any $X \in \mathbb{C}^{n\times m}$

$$\|\mathcal{A}\mathcal{S}(X) - b\|_2 = \|\mathcal{A}\mathcal{S}(X) - \mathcal{A}(D_\rho) - d\|_2$$
$$= \|\mathcal{A}\mathcal{S}(X) - \mathcal{A}(\mathcal{S}(D_0) + P) - d\|_2$$
$$= \|\mathcal{A}\mathcal{S}(X) - \mathcal{A}\mathcal{S}(D_0) - \mathcal{A}(P) - d\|_2.$$

Let us define $\tilde{\mathcal{A}} := \mathcal{A}\mathcal{S}$ and $\tilde{d} := \mathcal{A}(P) + d$. Thus, since $X^\sharp$ solves (3) then $X^\sharp$ also solves

$$\min_{X\in\mathbb{C}^{n\times m}} \|X\|_* \text{ subject to } \|\tilde{\mathcal{A}}(X) - \tilde{\mathcal{A}}(D_0) - \tilde{d}\|_2 \leq \tilde{\eta}, \tag{20}$$

because

$$\|\tilde{d}\|_2 = \|\mathcal{A}(P) + d\|_2 \leq \|\mathcal{A}(P)\|_2 + \|d\|_2$$
$$\leq \|\mathcal{A}\|_{F,2}\|P\|_F + \|d\|_2 \leq (\|\mathcal{A}\|_{F,2})\epsilon + \eta = \tilde{\eta},$$

the last inequality follows since we defined ϵ according to (7). We now apply all assumptions to (2) to conclude

$$\|D_0 - D^\sharp\|_F \leq \frac{c_1}{\sqrt{r}} \sum_{\ell=r+1}^{\min\{n,m\}} \sigma_\ell(D_0) + c_2\eta + c_3\epsilon, \tag{21}$$

where $c_1, c_2 > 0$ are constants depending only on $\delta_{2r}(\mathcal{A}\mathcal{S})$ from (2) and $c_3 = (\|\mathcal{A}\|_{F,2})c_2$.

□

D. Proof of Theorem II.3

Proof of Theorem II.3. Let $D \in \mathbb{C}^{n\times m}$ and recall that $\mathcal{S}(D) = \mathcal{N}\mathcal{F}^*(D) = \mathcal{N}(\check{D})$, so that we now think of this operator as a 2D non uniform Fourier transform, $\mathcal{N}$, acting on $\check{D} = \mathcal{F}^*(D) \in \mathbb{C}^{n\times m}$. By definition of $\mathcal{N}$, notice that the entry $\mathcal{N}(\check{D})(k,\ell)$ is obtained by a matrix product of $\check{D}$ with a rank 1 matrix, $\mathcal{M}^{k,\ell}$.

$$\begin{aligned}
&\mathcal{N}(\check{D})(k,\ell)\\
&= \sum_{u=1}^{n}\sum_{v=1}^{m} \check{D}(u,v)\mathbf{e}^{-\tilde{x}^{k,\ell}(u-\tilde{n}-1)}\mathbf{e}^{-\tilde{y}^{k,\ell}(v-\tilde{m}-1)}\\
&:= \langle \check{D}, \mathcal{M}^{k,\ell}\rangle.
\end{aligned}$$

Therefore, if we let $\mathcal{A}^p$ denote the $n\times m$ matrix whose $(k,\ell)\in[n]\times[m]$ entry is given by $\mathcal{A}(p,k,\ell)$, we have

$$\|\mathcal{A}\mathcal{S}(D)\|_2^2 = \frac{1}{N}\sum_{p=1}^{N}\langle \mathcal{A}^p, \mathcal{N}(\check{D})\rangle^2$$

where

$$\begin{aligned}
\langle \mathcal{A}^p, \mathcal{N}(\check{D})\rangle &= \sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathcal{A}(p,k,\ell)\mathcal{N}(\check{D})(k,\ell)\\
&= \sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathcal{A}(p,k,\ell)\langle \check{D}, \mathcal{M}^{k,\ell}\rangle\\
&= \sum_{u=1}^{n}\sum_{v=1}^{m}\check{D}(u,v)\sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathcal{A}(p,k,\ell)\mathcal{M}^{k,\ell}(u,v).
\end{aligned}$$

We now see that the term $\sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathcal{A}(p,k,\ell)\mathcal{M}^{k,\ell}(u,v)$ inherits a subgaussian distribution from $\mathcal{A}$. For example, from a generalization of Proposition 5.10 in [18] to the case of complex coefficients we may conclude

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathcal{A}(p,k,\ell)\mathcal{M}^{k,\ell}(u,v)\right| > \delta\right)\\
&\leq e\exp\left(-\frac{c\delta^2}{K^2\|\mathcal{M}(u,v)\|_F^2}\right) = e\exp\left(-\frac{c\delta^2}{nmK^2}\right)
\end{aligned}$$

where $c>0$ is an absolute constant (see [18]) and the term $\mathcal{M}(u,v)$ is used to denote the $n\times m$ matrix whose $(k,\ell)\in[n]\times[m]$ entry is given by $\mathcal{M}^{k,\ell}(u,v)$. The last equality holds since $\|\mathcal{M}(u,v)\|_F^2 = nm$. We now see that $\|\sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathcal{A}(p,k,\ell)\mathcal{M}^{k,\ell}(u,v)\|_{\psi_2} \leq K\sqrt{\frac{nm}{c}}$ for any $p\in[N]$ so that repeating the

argument with the sum $\sum_{u=1}^{n}\sum_{v=1}^{m}\check{D}(u,v)\sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathcal{A}(p,k,\ell)\mathcal{M}^{k,\ell}(u,v)$ we similarly obtain

$$\begin{aligned}&\mathbb{P}\left(\left|\sum_{v=1}^{n}\sum_{u=1}^{m}\check{D}(u,v)\sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathcal{A}(p,k,\ell)\mathcal{M}^{k,\ell}(u,v)\right|>\delta\right)\\&\leq e\exp\left(-\frac{c^2\delta^2}{K^2nm\|\check{D}\|_F^2}\right)=e\exp\left(-\frac{c^2\delta^2}{K^2\|D\|_F^2}\right),\end{aligned}$$

where the last equality holds since $\|\check{D}\|_F^2=\frac{1}{nm}\|D\|_F^2$.

Therefore each term $\langle\mathcal{A}^p,\mathcal{N}(\check{D})\rangle$ is subgaussian with subgaussian norm bounded by $\frac{K\|D\|_F}{c}$ so that by Lemma 5.14 in [18] each term $\langle\mathcal{A}^p,\mathcal{N}(\check{D})\rangle^2$ is subexponential with subexponential norm bounded by $\frac{2K^2\|D\|_F^2}{c^2}$.

We may apply an appropriate concentration inequality to the sum $\frac{1}{N}\sum_{p=1}^{N}\langle\mathcal{A}^p,\mathcal{N}(\check{D})\rangle^2$. For example, using Proposition 5.16 in [18] we have

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{N}\left|\sum_{p=1}^{N}\langle\mathcal{A}^p,\mathcal{N}(\check{D})\rangle^2-\mathbb{E}\langle\mathcal{A}^p,\mathcal{N}(\check{D})\rangle^2\right|>\delta\right)\\&\leq 2\exp\left(-\tilde{c}\min\left(\frac{Nc^4\delta^2}{4K^4\|D\|_F^4},\frac{Nc^2\delta}{2K^2\|D\|_F^2}\right)\right),\end{aligned}$$

where $\tilde{c}>0$ is an absolute constant (see [18]). We now consider $\mathbb{E}\langle\mathcal{A}^p,\mathcal{N}(\check{D})\rangle^2$.

$$\begin{aligned}&\mathbb{E}\langle\mathcal{A}^p,\mathcal{N}(\check{D})\rangle^2\\=&\mathbb{E}\left(\sum_{u,v}\check{D}(u,v)\sum_{k,\ell}\mathcal{A}(p,k,\ell)\mathbf{e}^{\tilde{x}^{k,\ell}(1+\tilde{n}-u)}\mathbf{e}^{\tilde{y}^{k,\ell}(1+\tilde{m}-v)}\right)^2\\=&\sum_{u,v,\tilde{u},\tilde{v}}\check{D}(u,v)\check{D}(\tilde{u},\tilde{v})\sum_{k,\ell}\mathbb{E}\left(\mathbf{e}^{\tilde{x}^{k,\ell}(\tilde{u}-u)}\right)\mathbb{E}\left(\mathbf{e}^{\tilde{y}^{k,\ell}(\tilde{v}-v)}\right)\\=&\sum_{u,v,\tilde{u},\tilde{v}}\check{D}(u,v)\check{D}(\tilde{u},\tilde{v})\mathbb{E}\left(\mathbf{e}^{\Delta(k,\ell)(\tilde{u}-u)}\right)\mathbb{E}\left(\mathbf{e}^{\Gamma(k,\ell)(\tilde{v}-v)}\right)\\&\cdot\left(\sum_{k=1}^{n}\sum_{\ell=1}^{m}\mathbf{e}^{\frac{k-1}{n}(\tilde{u}-u)}\mathbf{e}^{\frac{\ell-1}{m}(\tilde{v}-v)}\right)\\=&nm\sum_{u=1}^{n}\sum_{v=1}^{m}\check{D}(u,v)^2=\|D\|_F^2.\end{aligned}$$

The second equality follows by expanding the previous sum and using the zero-mean and unit variance assumptions of each $\mathcal{A}(p,k,\ell)$. The third equality is an abuse of notation to indicate that the sum over

k,ℓ does not depend on the terms $\mathbb{E}\left(\mathbf{e}^{\Delta(k,\ell)(\tilde{u}-u)}\right)\mathbb{E}\left(\mathbf{e}^{\Gamma(k,\ell)(\tilde{v}-v)}\right)$ since the entries of $\Delta,\Gamma\in\mathbb{R}^{n\times m}$ are i.i.d. and therefore the second to last equality follows by orthogonality of the complex exponentials, except in the case $u=\tilde{u}$ and $v=\tilde{v}$. To conclude, if we choose $\delta\leq\|D\|_F^2\leq\frac{2K^2\|D\|_F^2}{c^2}$, where the second inequality holds by assumption on K, we have shown

$$\mathbb{P}\left(\left|\|\mathcal{A}\mathcal{S}(D)\|_2^2-\|D\|_F^2\right|>\delta\right)\leq 2\exp\left(-\frac{\bar{c}N\delta^2}{4K^4\|D\|_F^4}\right),$$

for some absolute constant $\bar{c}=\tilde{c}c^4$. We now proceed to a covering number argument. Fix $\{u^k\}_{k=1}^r\subset\mathbb{C}^n,\{v^k\}_{k=1}^r\subset\mathbb{C}^m$, each a set of linearly independent vectors. We may provide $\mathcal{D}_r=\{X\in\mathbb{C}^{n\times m}:\|X\|_F\leq 1,X=\sum_{k=1}^r\sigma_k(X)u^kv^{k*}\}$ with a radius-γ Frobenius norm covering, $\tilde{\mathcal{D}}_r$, with covering number not exceeding $\left(1+\frac{6}{\gamma}\right)^{r(n+m+1)}$ (see for example [15]). Therefore,

$$\begin{aligned}
&\mathbb{P}\left(\forall\tilde{D}\in\tilde{\mathcal{D}}_r\ \mid\ \left|\|\mathcal{A}\mathcal{S}(\tilde{D})\|_2^2-\|\tilde{D}\|_F^2\right|>\delta\right)\\
&\leq 2\exp\left(-\frac{\bar{c}N\delta^2}{4K^4\|\tilde{D}\|_F^4}\right)\left(1+\frac{6}{\gamma}\right)^{r(n+m+1)}\\
&\leq 2\exp\left(r(n+m+1)\log\left(1+\frac{6}{\gamma}\right)-\frac{\bar{c}N\delta^2}{4K^4}\right).
\end{aligned}$$

Thus, for any $D\in\mathcal{D}_r$ there is some $\tilde{D}\in\tilde{\mathcal{D}}_r$ with $\|D-\tilde{D}\|_F\leq\gamma$ so that if we denote $\mathcal{B}=(\mathcal{A}\mathcal{S})^*\mathcal{A}\mathcal{S}-\mathcal{I}$, where $\mathcal{I}:\mathbb{C}^{n\times m}\mapsto\mathbb{C}^{n\times m}$ is the identity operator, then

$$\begin{aligned}
&\left|\|\mathcal{A}\mathcal{S}(D)\|_2^2-\|D\|_F^2\right|=|\langle\mathcal{B}(D),D\rangle|\\
=&\left|\langle\mathcal{B}(\tilde{D}),\tilde{D}\rangle+\langle\mathcal{B}(D+\tilde{D}),D-\tilde{D}\rangle\right|\\
\leq&\left|\langle\mathcal{B}(\tilde{D}),\tilde{D}\rangle\right|+\left|\langle\mathcal{B}(D+\tilde{D}),D-\tilde{D}\rangle\right|\\
\leq&\delta+\|\mathcal{B}(D+\tilde{D})\|_F\|D-\tilde{D}\|_F,
\end{aligned}$$

with probability according to our previous derivation. Letting $\|\mathcal{B}\|_r:=\max_{X\in\mathcal{D}_r}\|\mathcal{B}(X)\|_F$, since $\mathcal{B}$ is a Hermitian operator we may take the maximum over all $D\in\mathcal{D}_r$ to conclude

$$\max_{D\in\mathcal{D}_r}\left|\|\mathcal{A}\mathcal{S}(D)\|_2^2-\|D\|_F^2\right|=\|\mathcal{B}\|_r\leq\frac{\delta}{1-2\gamma}.$$

This ends the proof.

□