Completion of Structured Low-Rank Matrices via Iteratively Reweighted Least Squares

Christian Kümmerle
Department of Mathematics
Technische Universität München
Email: c.kuemmerle@tum.de

Claudio M. Verdun
Department of Mathematics
Technische Universität München
Email: verdun@ma.tum.de

***Abstract*—We propose a new Iteratively Reweighted Least Squares (IRLS) algorithm for the problem of completing a low-rank matrix that is *linearly structured*, e.g., that possesses a Hankel, Toeplitz or block-Hankel/Toeplitz structures, which is of relevance for the harmonic retrieval or super-resolution problem. The algorithm optimizes a non-convex surrogate of the rank by minimizing well-chosen quadratic upper bounds of the smoothed surrogate.**

We establish a *quadratic local convergence rate* of the developed IRLS strategy if the linear structure is *Hankel*, with high probability if the provided entries of the matrix are sampled uniformly at random, and if the matrix to be completed fulfills a suitable coherence condition. Our strategy combines *computational efficiency*, as the dimensionality of its optimization variables scales sub-linearly in the matrix dimensions, with a favorable *data efficiency*, as it can be observed in experiments on hard completion tasks.

In particular, our experiments show that the proposed algorithm exhibits an empirical recovery probability close to one from fewer samples than existing state-of-the-art approaches for the Hankel matrix completion task arising from the problem of spectral super-resolution of frequencies with small separation.

I. Introduction

It is well known that a number of seemingly unrelated problems as *phase retrieval* in x-ray crystallography and ptychography [1], [2], *blind deconvolution* in wireless communications [3] and *data completion* in recommender systems [4] can formulated as problems of identifying and recovering a low-rank matrix from underdetermined linear measurements or observations. One benefit of these formulations is that it is possible to prove that certain tractable algorithms solve the underlying problems [5]–[7] if an optimal or near-optimal amount of samples is provided to the algorithms, and if the provided samples are generic or random enough.

However, for some problems in parallel MRI [8], system identification [9], [10], direction of arrival [11] and the interpolation of seismic data [12], the setup allows us to exploit *an additional linear structure* of the low-rank matrices to be identified. In particular, in these applications, it is possible to equip a low-rank matrix recovery algorithm with prior knowledge about the solution such that a low-dimensional linear subspace is provided in which the target (approximately) low-rank matrix has to lie. By doing this, the number of samples to ensure exact recovery can be further reduced [13].

To understand how linearly structured low-rank matrices arise in the above domains, consider a one-dimensional discrete signal $x \in \mathbb{C}^n$ that is a superposition of $r \ll n$ exponentials, i. e.,

$$x_t = \sum_{k=1}^{r} \alpha_k \mathrm{e}^{2\pi i f_k (t-1)} \tag{1}$$

for $t \in \{1, \dots, n\}$, where $\alpha_1, \dots, \alpha_r \in \mathbb{C}$ are complex amplitudes and $f_1, \dots, f_r \in [0, 1)$ are fixed frequencies. The problem of learning the model parameters $\{\alpha_k\}_{k=1}^r$ and $\{f_k\}_{k=1}^r$ from incomplete or noisy samples of x is often also called super-resolution [14], harmonic retrieval [15], line spectral estimation [16] or spectral compressed sensing [17] and is a classical problem in signal processing. If we now fix d_1 such that $r \leqslant d_1 \leqslant n - r + 1$, define $d_2 = n - d_1 + 1$ and consider the linear *Hankel operator* $\mathcal{H} : \mathbb{C}^n \to \mathbb{C}^{d_1 \times d_2}, x \mapsto \mathcal{H}(x) = (\mathcal{H}(x))_{i,j=1}^{d_1,d_2} = (x_{i+j-1})_{i,j=1}^{d_1,d_2}$, it can be shown that the Hankel matrix $\mathcal{H}(x)$ associated to a spectrally sparse signal x as in (1) has rank r [18], or to put it differently, $X = \mathcal{H}(x)$ is a matrix that fulfills both $\operatorname{rank}(X) = r$ and $X \in \operatorname{Ran} \mathcal{H}$.

In the case of missing or noisy samples in the vector x, this double structure can be used to identify the model parameters of (1) nevertheless.

Therefore, in this paper, we are interested in the following *structured low-rank matrix completion problem*: Let $m < n$, $\Omega \subset \{1, \dots, n\}$ be an index set of cardinality $|\Omega| = m$, and $\mathcal{P}_\Omega : \mathbb{C}^n \to \mathbb{C}^m, x \mapsto \mathcal{P}_\Omega = (x_t)_{t \in \Omega}$ its subsampling operator, and let $P_\Omega(x)$ be the partial samples of a discrete signal x. We consider

$$\widehat{Z} = \mathcal{H}(\widehat{z}) \text{ s.t. } \widehat{z} = \operatorname*{arg\,min}_{z \in \mathbb{C}^n, \mathcal{P}_\Omega(z) = \mathcal{P}_\Omega(x)} \operatorname{rank}(\mathcal{H}(z)), \tag{2}$$

the task of finding a linearly structured matrix $\mathcal{H}(\widehat{z})$ of low rank that is also compatible with the samples $\mathcal{P}_\Omega(x)$.

Related to this is the *structured low-rank approximation problem*

$$\widehat{z} = \operatorname*{arg\,min}_{\substack{z \in \mathbb{C}^n \text{ s.t.} \\ \operatorname{rank}(\mathcal{H}(z)) \leqslant r}} \|\mathcal{H}(z) - \mathcal{H}(x)\|_{F(w)}^2, \tag{3}$$

where $\|\cdot\|_{F(w)}$ is a suitable weighted Frobenius norm [9], [19] and $r < \min(d_1, d_2)$ is fixed.

Due to the interplay of linear subspace and non-convex low-rank structures, finding efficient computational approaches to approximately or exactly solve (2) and (3) is challenging.

A classical approach to solve the problems is based on alternating projections [19], [20] (sometimes called *Cadzow's method*), as the projections onto the rank-r manifold and the linear subspace parametrized by $\mathcal{H}$ can be efficiently calculated, but this approach lacks provable algorithmic guarantees. A more recent approach based on *nuclear norm minimization* (NNM) [8], [13] has been shown to shown to solve the completion problem given a near-optimal amount of samples, but this has its computational limitations as it is equivalent to solving a semidefinite program (SDP) with an order of at least n^2 variables. Computationally more advantageous non-convex optimization strategies such as *iterative hard thresholding* (IHT) [21] and *projected gradient descent* (PGD) [22] have also been developed. Despite this progress which also involves algorithmic guarantees, it is hard to quantify which algorithmic approach provides an optimal tradeoff between *data efficiency*, i. e., the ability to identify the model for a number of samples m as small as possible, *space complexity* and *time complexity*.

Computational challenges arise in particular for real-world engineering instances of the problem, where n is very large and a quadratic growth of dimensionalty from x to $\mathcal{H}(x)$ can be a computational issue, if $\mathcal{H}$ is the Hankel operator. Furthermore, for multi-dimensional harmonic retrieval problems [8], [23], $\mathcal{H}$ can be a block Hankel matrix operator whose size is non-trivial except of for very small problem instances.

In this work, we develop a new Iteratively Reweighted Least Squares (IRLS) algorithm tailored to the structured low-rank matrix estimation problems (2) and (3), and provide a local convergence analysis. This continues a line of work on algorithms of this kind for the recovery of unstructured low-rank matrices [24]–[26].

II. Our Approach

To derive optimization-based approaches for the problems (2) and (3), the idea is to replace the non-convex and non-smooth rank by a smooth, but still non-convex surrogate whose minimizers are closely related to the minimizers of the rank. The first contribution in this direction was done by [27], [28] and their idea was to use $\log\det(X+\epsilon\mathrm{Id})$ as a smooth surrogate for $\mathrm{rank}(X)$ in the case of positive semidefinite matrices, where ϵ is some suitable regularization parameter. Inspired by this strategy, we propose a more suitable *logdet* based objective $\mathcal{J}$ that can be used for general non-square matrices and comes with computational advantages. For the completion problem, which is the focus of this paper, it is defined as

$$\mathcal{J}(x,\epsilon) := \sum_{i=1}^{R} \log(\sigma_i(\mathcal{H}(x))^2+\epsilon^2) + \sum_{i=R+1}^{\min(d_1,d_2)} \log(\epsilon^2) + \tfrac{\sigma_i^2(\mathcal{H}(x))}{\epsilon^2},$$

where R is an upper bound for the target rank r.

In the following, we propose an algorithm that can be seen as a sequential minimization of local quadratic upper bounds on the functional $\mathcal{J}(x,\epsilon)$ under a data constraint. The precise formulation of the proposed algorithm `StrucHMIRLS` can be found in Algorithm 1.

Algorithm 1: `Structured Harmonic Mean Iteratively Reweighted Least Squares`

Input: Operators $\mathcal{P}_\Omega : \mathbb{C}^n \to \mathbb{C}^m$ and $\mathcal{H} : \mathbb{C} \to \mathbb{C}^{d_1\times d_2}$, data vector $y \in \mathbb{R}^m$, rank parameter r, rank estimate $R > r$

Output: Sequence $(x^{(k)})_{k\geqslant 1} \subset \mathbb{C}^n$.

Initialize $\epsilon_0 = \sigma_1(\mathcal{H}(\mathcal{P}_\Omega^*(y)))$, $W^{(0)} = \epsilon_0^2\mathrm{Id} \in \mathbb{C}^{n\times n}$.

for $k = 1, 2, \dots$ **do**

$$x^{(k)} = \operatorname*{arg\,min}_{z\in\mathbb{C}^n, \mathcal{P}_\Omega(z)=y} \langle z, W^{(k-1)} z\rangle_{\ell_2}, \quad (4)$$

$$\mathcal{T}_R(\mathcal{H}(x^{(k)})) = U_R^{(k)}\mathrm{diag}(\sigma_i^{(k)})V_R^{(k)*}, \quad (5)$$

$$\epsilon_k = \min(\epsilon_{k-1}, \sigma_{r+1}^{(k)}), \quad (6)$$

$$W^{(k)} = \mathcal{H}^*\widetilde{W}^{(k)}\mathcal{H}, \quad (7)$$

where $\mathcal{T}_R$ computes the the best rank-R approximation of a matrix and $\widetilde{W}^{(k)}$ is defined as in (8).

In the algorithm, (4) corresponds to solving an $n\times n$ linear system coming from the linearly constrained least squares problem, and the updates (6) and (7) of the smoothing parameter ϵ_k and the weight matrix $W^{(k)}$ can be defined using only information from the best rank-R approximation $\mathcal{T}_R(\mathcal{H}(x^{(k)}))$ of the current matrix of (5). More precisely, for each iteration k, the action of $\widetilde{W}$ on a matrix X is given by

$$\begin{aligned}\widetilde{W}^{(k)}(X) = {} & U_R^{(k)}[H^{(k)} \circ (U_R^{(k)*} X V_R^{(k)})]V_R^{(k)*} \\ & + U_R^{(k)} D^{(k)*} U_R^{(k)*} X (Id - V_R^{(k)} V_R^{(k)*}) \\ & + (Id - U_R^{(k)} U_R^{(k)*}) X V_R^{(k)} D^{(k)*} V_R^{(k)*} \\ & + \epsilon_k^{-2}(Id - U_R^{(k)} U_R^{(k)*}) X (Id - V_R^{(k)} V_R^{(k)*}),\end{aligned} \quad (8)$$

where $A\circ B$ denotes the entrywise product of the matrices A and B, $U_R^{(k)}, V_R^{(k)}$ denote the matrices consisting of the R first left and right singular vectors of $\mathcal{H}(x^{(k)})$, respectively, and where the matrices $H^{(k)}, D^{(k)} \in \mathbb{C}^{R\times R}$ are given by

$$H_{ij}^{(k)} = 2\left[\sigma_i^2(\mathcal{H}(x^{(k)})) + \sigma_j^2(\mathcal{H}(x^{(k)})) + 2\epsilon_k^2\right]^{-1}, \text{ for } i,j \leqslant R,$$

$$D^{(k)} = \mathrm{diag}\left(\left\{2\left[\sigma_j^2(\mathcal{H}(x^{(k)})) + 2\epsilon_k^2\right]^{-1}\right\}_{j=1}^{R}\right).$$

For the design of the algorithm, it is crucial that the quadratic upper bounds on $\mathcal{J}(x^{(k)},\epsilon_k)$ are well-chosen, since the optimization landscape of $\mathcal{J}$ is in general extremely non-convex. The precise shape of these upper bounds is encoded in the weight matrix choice (7) and (8), and our choice is inspired by the weight matrix rule of [25], where a *harmonic mean* of certain previous weight matrices rules [24], [26] was considered for the unstructured low-rank matrix recovery problem to optimize a Schatten-p objective.

An important difference to the rule of [25] is that due to the low-rank structure of (8), it is sufficient to work with *partial SVD* information from a best rank-R approximation of the iteration matrices $\mathcal{H}(x^{(k)})$.

We note that our algorithm has strong connections to Newton methods modified to minimize non-convex functions [29], which can be seen by relating $\widetilde{W}$ to the Hessian of the objective $\mathcal{J}$.

Furthermore, we note that while an IRLS algorithm for structured low-rank matrix recovery has been already proposed in [30], the algorithm of [30] does not optimize a rank surrogate of the Hankel matrix $\mathcal{H}(x)$ itself, but of a half-circulant extension thereof. This half-circulant extension is not expected to be very low-rank even if $\mathcal{H}(x)$ is, which renders the connection to the original problem unclear.

III. Computational Aspects

In practice, if $\mathcal{H}$ is, for example, a Hankel operator, the best rank-R approximations of (5) can be computed very efficiently by using that Hankel matrices are restricted circulant matrices, which allow for the usage FFTs due to the diagonalization of circulant matrices by the discrete Fourier transform.

The linear system of (4) can be solved by an iterative solvers as a conjugate gradients method, since the corresponding system matrix is positive definite. In the Hankel case, iterative solvers for (4) benefit from fast matrix-vector multiplications of the system matrix. This property is illustrated by the following lemma.

Lemma 1. *If $\mathcal{H} : \mathbb{C}^n \to \mathbb{C}^{d_1\times d_2}$ is the Hankel operator, the multiplication of $W^{(k)} \in \mathbb{C}^{n\times n}$ with a vector $v \in \mathbb{C}^n$ can be computed in $O(nR^2 + nR\log n)$ operations.*

Thus, if a tight upper estimate R of the true rank r is provided and $r \ll n$, matrix-vector multiplication with $W^{(k)}$ is basically linear in the dimensionality of x.

We note that the rule (6) for the smoothing parameter ϵ_k has the disadvantage of necessitating the true model order r. Different rules that also allow for a (more involved) theoretical analysis, but which do not use the knowledge of r, are also possible [31], [32].

IV. Theoretical Results

In the case that $\mathcal{H}$ is the Hankel operator, we establish in the following a result about *local convergence with quadratic rate* of `StrucHMIRLS` starting from an iterate $\mathcal{H}(x^{(k_*)})$ that is close enough to a low-rank Hankel matrix $X_0 = \mathcal{H}(x_0)$.

We say that a rank-r matrix X with singular value decomposition $X = U\Sigma V^*$ is μ_0-*incoherent* if

$$\max_{1\leqslant i\leqslant d_1} \|U^* e_i\|_2 \leqslant \sqrt{\frac{\mu_0 r}{d_1}}, \quad \max_{1\leqslant j\leqslant d_2} \|V^* e_j\|_2 \quad \leqslant \sqrt{\frac{\mu_0 r}{d_2}}, \quad (9)$$

where e_i denotes the vector that is different from zero only at the i-th entry.

Theorem 1. *Let $x_0 \in \mathbb{C}^n$ be vector such that $\mathcal{H}(x_0) \in \mathbb{C}^{d_1\times d_2}$ is a rank-r matrix that is μ_0-incoherent. Let $\Omega \subset \{1,\dots,n\}$ be a collection of m indices sampled i.i.d. uniformly with replacement. Assume $n \geqslant 9$ and that*

$$\|\mathcal{H}(x^{(k_*)}) - \mathcal{H}(x_0)\| \leqslant \zeta\sigma_r(\mathcal{H}(x_0))$$

for some $\zeta < \min\left(\frac{1}{2}, \frac{1}{21}\sqrt{\frac{m}{4n\log^2(n)}}\right)$. Then, if

$$m \geqslant 128\beta\mu_0 r\log(n), \quad (10)$$

there is a constant $c(n,r,\kappa)$ such that with probability at least $1-n^{-2}-d_2^{2-2\beta}$,

$$\|\mathcal{H}(x^{(k+1)}) - \mathcal{H}(x_0)\| \leqslant \frac{c(n,r,\kappa)}{\sigma_r(\mathcal{H}(x_0))}\|\mathcal{H}(x^{(k)}) - \mathcal{H}(x_0)\|^2, \quad \forall k \geqslant k_*.$$

Here, $\kappa = \sigma_1(\mathcal{H}(x_0))/\sigma_r(\mathcal{H}(x_0))$ is the condition number of $\mathcal{H}(x_0)$ and $c(n,r,\kappa) = 4(\frac{16}{9}n^2+1)r(3+4\kappa)$.

It is instructive to compare the sufficient condition for the sample complexity (10) with corresponding conditions for local convergence results for PGD [22] and IHT [21], which require $\Omega(\mu_0^2\kappa^2 r^2\log(n))$ and $\Omega(\mu_0\kappa^6 r^2\log(n))$ samples respectively. We note that a *global convergence* theory for NNM [13] is available if $\Omega(\mu_0 r\log^4(n))$ samples are provided.

Moreover, it is important to highlight the contribution of this result. As noted in [25], certain sufficient conditions for low-rank matrix recovery, such as *restricted isometry properties* or *null space properties*, do not hold for the matrix completion case. Thus, with Theorem 1, we solve an open question about local convergence of IRLS strategies for entrywise measurements, inspired by [21], [33], [34], for *incoherent* matrices. While this is not the focus of the current paper, we note that it is possible to generalize our analysis of local convergence to appropriate IRLS algorithms for unstructured low-rank matrix completion problems, as the linear (Hankel) structure does not play a prominent role in our arguments.

V. Proof of Convergence

In order to establish Theorem 1, we first state a lemma about the spectral norm $\|\cdot\|$ of the sampling operator $\mathcal{P}_\Omega$.

Lemma 2 ([33, Proposition 5]). *With probability at least $1-n^{2-2\beta}$, the maximum number of repetitions of any entry in Ω is less than $\frac{8}{3}\beta\log(n)$ for $n \geqslant 9$ and $\beta > 1$. Consequently, with high probability we have $\|\mathcal{P}_\Omega\| \leqslant \frac{8}{3}\beta\log(n)$.*

In order to study the convergence of the proposed algorithm, some notation regarding Hankel matrices must be established. The adjoint of $\mathcal{H}$ is denoted by $\mathcal{H}^* : \mathbb{C}^{d_1\times d_2} \to \mathbb{C}^n$ and its action is given by $\mathcal{H}^*(X) = \{\sum_{i+j=a} X_{ij}\}_{a=0}^{n-1}$. We will define the operator $\mathcal{D}^2 = \mathcal{H}^*\mathcal{H}$ and $\mathcal{G} = \mathcal{H}\mathcal{D}^{-1}$. Then the adjoint of $\mathcal{G}$ is given by $\mathcal{G}^* = \mathcal{D}^{-1}\mathcal{H}^*$.

Besides that, the tangent space T of the embedded rank r matrix manifold at the point $\mathcal{H}(x)$ is defined as

$$T := \{UA^* + BV^* : \ A\in\mathbb{C}^{d_2\times r}, \ B\in\mathbb{C}^{d_1\times r}\}. \quad (11)$$

The orthogonal projections onto the subspaces spanned by U, V and T as well as the its orthogonal complement of the later will be respectively denoted by $\mathcal{P}_U$, $\mathcal{P}_V$, $\mathcal{P}_T$ and $\mathcal{P}_{T^\perp}$, respectively.

Lemma 3 ([13, Lemma 3]). *Let $U \in \mathbb{C}^{d_1\times r}$ and $V \in \mathbb{C}^{d_2\times r}$ be two μ_0-incoherent orthogonal matrices. Let T be the tangent space defined in* (11) *and define $p^{-1} = n/m$ and $c_s = \max\left\{\frac{n}{d_1}, \frac{n}{d_2}\right\}$. Then*

$$\|\mathcal{P}_T\mathcal{G}\mathcal{G}^*\mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{G}\mathcal{P}_\Omega\mathcal{G}^*\mathcal{P}_T\| \leqslant \sqrt{\frac{32\mu c_s r\log(n)}{m}} \quad (12)$$

holds with probability at least $1-n^{-2}$ provided that

$$m \geqslant 128\mu c_s r\log(n).$$

Lemma 3 can be seen as a *restricted isometry property* (RIP) on the tangent space (see [5, Section 4.2] for some discussion). As a consequence of it, it is possible to establish a crucial inequality that controls the projection onto the tangent space.

Lemma 4. *Let $x \in \mathbb{C}^n$ and T be the tangent space of the embedded rank r matrix manifold at the point $\mathcal{H}(x)$. Assume that* (12) *holds. Then, for any $\mathcal{H}(x) = \mathcal{D}\mathcal{G}x$, it holds that*

$$\|\mathcal{P}_T\mathcal{H}(x)\|_F < 3n\|\mathcal{P}_{T^\perp}\mathcal{H}(x)\|_F \quad \forall x \in ker(\mathcal{P}_\Omega).$$

Inspired by the Riemannian optimization algorithms for low-rank matrix completion [34], [35], the key ingredient for the proof is a perturbation argument on the tangent space, showing that if two matrices are sufficiently close, then we can transfer the properties showed in the Lemmas 3 and 4 from the matrix of the previous iteration to the matrix of the next iteration of the algorithm. More precisely, let $x^{(k)}$ be the current iterate of the algorithm, $\mathcal{T}_r(\mathcal{H}(x^{(k)})) = U_r^{(k)}\Sigma_k V_r^{(k)*}$ and T_k be the tangent space of the r-rank matrix manifold at $\mathcal{H}(x^{(k)})$, that is, $T_k = \{U_r^{(k)}A^* + BV_r^{(k)*} : \ A\in\mathbb{C}^{d_2\times r}, \ B\in\mathbb{C}^{d_1\times r}\}$.

Lemma 5 ([21, Lemma 8]). *Assume that $0 < \varepsilon_0 < 1/10$ and that the following three conditions hold:*

- $\|\mathcal{P}_\Omega\| \leqslant \frac{8}{3}\log(n)$,
- $\|\mathcal{P}_T\mathcal{G}\mathcal{G}^*\mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{G}\mathcal{P}_\Omega\mathcal{G}^*\mathcal{P}_T\| \leqslant \varepsilon_0$
- $\frac{\|\mathcal{H}(x^l)-\mathcal{H}(x_0)\|_F}{\sigma_{\min}(\mathcal{H}(x_0))} \leqslant \frac{1}{21}\sqrt{\frac{m}{4n\log^2(n)}}$.

Then we have

$$\|\mathcal{P}_{T_l}\mathcal{G}\mathcal{G}^*\mathcal{P}_{T_l} - p^{-1}\mathcal{P}_{T_l}\mathcal{G}\mathcal{P}_\Omega\mathcal{G}^*\mathcal{P}_{T_l}\| \leqslant 4\varepsilon_0,$$
$$\|\mathcal{P}_{T_l}\mathcal{H}(x)\|_F < \frac{4n}{3}\|\mathcal{P}_{T_l^\perp}\mathcal{H}(x)\|_F \quad \forall x \in ker(\mathcal{P}_\Omega). \quad (13)$$

We can finally sketch the proof of our main result:

Proof of Theorem 1. Let $\eta^{(k)} = \mathcal{H}(x^{(k)} - x_0)$ for $k \in \mathbb{N}$. Define T_k as the space tangent of the manifold of rank-r matrices at $\mathcal{H}(x^{(k)})$. Using Lemma 5 for T_k, it follows that

$$\begin{aligned}\left\|\eta^{(k+1)}\right\|^2 &\leqslant \left\|\eta^{(k+1)}\right\|_F^2 = \left\|P_{T_k}\eta^{(k+1)}\right\|_F^2 + \left\|P_{T_k^\perp}\eta^{(k+1)}\right\|_F^2 \\ &\leqslant \left(\frac{16n^2}{9}+1\right)\left\|P_{T_k^\perp}\eta^{(k+1)}\right\|_F^2.\end{aligned} \quad (14)$$

Using the calculations of [25, Lemma 20], we obtain that

$$\begin{aligned}\left\|P_{T_k^\perp}\eta^{(k+1)}\right\|_F^2 &= \left(\sigma_{r+1}^2(\mathcal{H}(x^{(k)})) + \epsilon_k^2\right)\left\|\eta_{T_k^\perp}^{(k+1)}\right\|_{F(W^{(k)})}^2 \\ &\leqslant \left(\sigma_{r+1}^2(\mathcal{H}(x^{(k)})) + \epsilon_k^2\right)\left\|\eta^{(k+1)}\right\|_{F(W^{(k)})}^2,\end{aligned} \quad (15)$$

where $\|X\|_{F(\widetilde{W}^{(k)})} := \sqrt{\mathrm{Tr}\left(X * \widetilde{W}^{(k)}(X)\right)}$ is the weighted Frobenius norm of a matrix X with respect to $\widetilde{W}^{(k)}(X)$, defined as in (8). Noting that $x^{(k+1)}$ is the solution of the constrained least-squares problem with weight operator $W^{(k)}$, as described in (7), we see that (cf. [25, Lemma 16])

$$0 = \langle \widetilde{W}^{(k)} \mathcal{H}(x^{(k+1)}), \eta^{(k+1)} \rangle = \langle \widetilde{W}^{(k)}(\eta^{(k+1)} + \mathcal{H}(x_0)), \eta^{(k+1)} \rangle,$$

which is equivalent to

$$\begin{aligned} \left\|\eta^{(k+1)}\right\|^2_{F(W^{(k)})} &= \langle \widetilde{W}^{(k)} \eta^{(k+1)}, \eta^{(k+1)} \rangle \\ &= -\langle \widetilde{W}^{(k)}(\mathcal{H}(x_0)), \eta^{(k+1)} \rangle \leqslant \left\|\widetilde{W}^{(k)}(\mathcal{H}(x_0))\right\|_{S_1} \left\|\eta^{(k+1)}\right\| \end{aligned} \tag{16}$$

Here $\|\cdot\|_{S_1}$ denotes the Schatten-1 (or nuclear) norm. Using the notation $X_0 = \mathcal{H}(x_0)$, we proceed as in the proof of [25, Lemma 21], and we estimate

$$\begin{aligned} \left\|\widetilde{W}^{(k)}(X_0)\right\|_{S_1} &\leqslant \left\|U_r^{(k)}[(H^{(k)})^r_{i,j=1} \circ (U_r^{(k)*} X_0 V_r^{(k)})] V_r^{(k)*}\right\|_{S_1} \\ &+ 2\sigma_r(k)^{-2} \left\|U_r^{(k)*} X_0 (Id - V_r^{(k)} V_r^{(k)*})\right\|_{S_1} \\ &+ 2\sigma_r(k)^{-2} \left\|(Id - U_r^{(k)} U_r^{(k)*}) X_0 V_r^{(k)}\right\|_{S_1} \\ &+ \epsilon_k^{-2} \left\|(Id - U_r^{(k)} U_r^{(k)*}) X_0 (Id - V_r^{(k)} V_r^{(k)*})\right\|_{S_1} \\ &\leqslant \frac{r}{\sigma_r(X_0)(1-\zeta)^2} \left[1 + 4\zeta + 2\kappa \frac{\|\eta^{(k)}\|^2}{\epsilon_k^2}\right]. \end{aligned} \tag{17}$$

We now proceed iteratively. Using that $\zeta < 1/2$ and combining inequalities (14), (15), (16) and (17), we obtain

$$\begin{aligned} \left\|\eta^{(k+1)}\right\| &\leqslant \frac{4\big(16n^2+9\big)\big(\sigma^2_{r+1}(\mathcal{H}(x^{(k)})) + \epsilon_k^2\big) r}{9\sigma_r(X_0)} \left[3 + \frac{2\kappa\|\eta^{(k)}\|^2}{\epsilon_k^2}\right] \\ &= \frac{4(16n^2+9)r}{9\sigma_r(X_0)} \left[3\big(\sigma^2_{r+1}(\mathcal{H}(x^{(k)}) + \epsilon_k^2\big) + 4\kappa\|\eta^{(k)}\|^2\right] \\ &\leqslant \frac{4(16n^2+9)r}{9\sigma_r(X_0)} \left[3 + 4\kappa\right] \|\eta^{(k)}\|^2, \end{aligned} \tag{18}$$

where we used in the last line that $\sigma_{r+1}(\mathcal{H}(x^{(n)})) = \|\mathcal{H}(x^{(k)}) - \mathcal{T}_r(\mathcal{H}(x^{(k)}))\| \leqslant \|\mathcal{H}(x^{(k)}) - X_0\| = \|\eta^{(k)}\|$ and $\epsilon_k \leqslant \sigma_{r+1}(\mathcal{H}(x^{(k)}))$, which holds due to its definition. Finally, we have established the local quadratic convergence with constant $c(n,r,k) = 4(\frac{16}{9}n^2+1)r(3+4\kappa)$.

□

VI. Numerical Results

In a first experiment, we consider the completion of Hankel matrices $\mathcal{H}(x)$ from m sample coordinates T of $x = (x(0), \ldots, x(n-1))$ that are drawn uniformly at random, $n = 127$ and $x(t) = \sum_{k=1}^r \alpha_k e^{(2\pi i f_k)t}$, where the f_k and c_k are sampled independently such that $f_k \sim \mathcal{U}([0,1]), |\alpha_k| = 1 + 10^{c_k}$, $c_k \sim \mathcal{U}([0,1])$, as in [22]. In Figure 1, the empirical recovery probabilities averaged over 50 simulations for each pair of m and r are documented in comparison with those of algorithms [13], [21], [22], [36], [37]. We observe that `StrucHMIRLS` exhibits the best performance, with successful recovery already when $m \approx 2r$, despite the fact that some frequencies f_k will be very close if r is not too small, which compromises the performance of, e.g., atomic norm minimization [38], [39].

Furthermore, we use a denoising variant of `StrucHMIRLS` tailored to the problem (3) in a second experiment, in order to investigate its performance for frequency estimation under the presence of additive Gaussian noise on equispaced samples from a signal that is a sum of two frequencies located at $f_1 = 0.35$ and $f_2 = 0.40$ (both with unitary amplitude), following Section VI of [40].

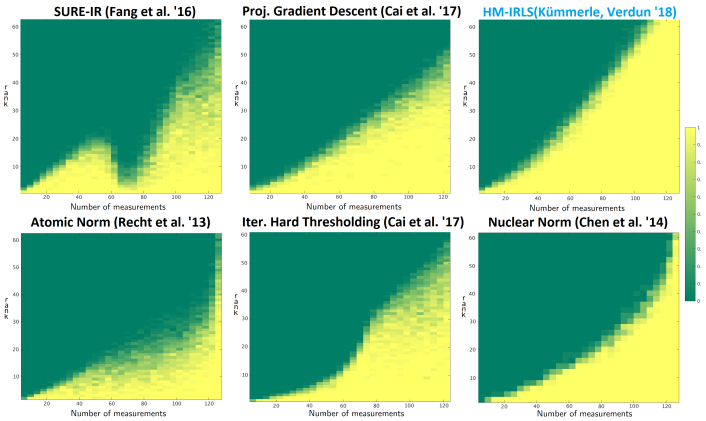

Fig. 1: Hankel matrix completion, m measurements of vector $x \in \mathbb{C}^n$ with $n = 123$. x-axis: number of measurements m, y-axis: model order r.

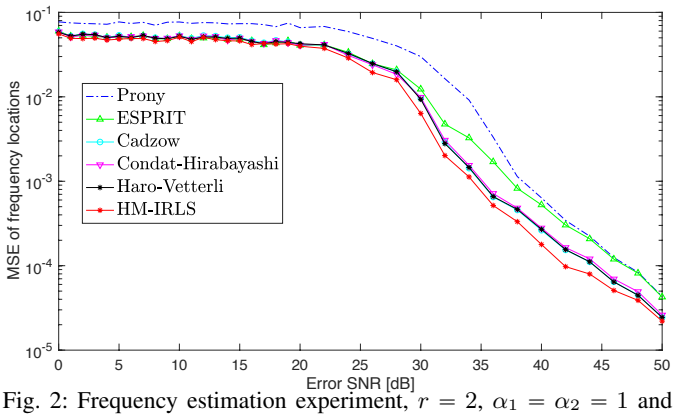

Fig. 2: Frequency estimation experiment, $r = 2$, $\alpha_1 = \alpha_2 = 1$ and $f_1 = 0.35$ and $f_2 = 0.40$.

After denoising, we use ESPRIT to obtain the frequencies, which is arguably one of best algorithms for frequency estimation for low noise levels [41]. As a comparison, we use the algorithms [19], [20], [40] (combined with ESPRIT for frequency retrieval, respectively), vanilla-ESPRIT [42] and Prony's method. For our method, we choose the regularization parameter λ according to an adaptive rule that uses the information of the model order $r = 2$.

The results corresponding to an average over 500 independent noise realizations for each SNR value can be seen in Figure 2, and our method consistently obtains a lower MSE on the vector of frequencies $f = (f_1, f_2)$ than the competing methods across different noise SNRs.

VII. Conclusion

In this paper, we formulate an IRLS algorithm suited for structured low-rank matrix recovery problems. We exhibited beneficial computational properties and also theoretical guarantees for local convergence. The numerical experiments suggest a very competitive statistical accuracy compared to other state-of-the-art methods for the completion problem. They also indicate that the algorithm can be used as a preprocessing step in harmonic retrieval problems. Further developments are the establishment of a global convergence theory

and extensive experiments with large real data such as those from model order reduction in control theory.

VIII. Acknowledgements

C.K. acknowledges the support of the German-Israeli Foundation (GIF) through the grant G-1266-304.6/2015. C.M.V. acknowledges the support of the German Research Foundation (DFG) through the Grant PO 1347/3-1.

References

[1] E. J. Candès, Y. Eldar, T. Strohmer, and V. Voroninski, "Phase Retrieval via Matrix Completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 199–225, 2013.

[2] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[3] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1711–1732, 2014.

[4] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[5] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[6] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 608–622, 2016.

[7] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *arXiv preprint arXiv:1809.09573*, 2018.

[8] K. H. Jin, D. Lee, and J. C. Ye, "A general framework for compressed sensing and parallel mri using annihilating filter based low-rank hankel matrix," *IEEE Trans. Comput. Imaging*, vol. 2, no. 4, pp. 480–495, 2016.

[9] I. Markovsky, *Structured low-rank approximation: Algorithms, Implementation, Applications*, 2nd ed. Springer International Publishing, 2019. [Online]. Available: https://www.springer.com/la/book/9783319896199

[10] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 3, pp. 1235–1256, 2009.

[11] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing*. Academic Press, 2017, vol. 7, pp. 509–581.

[12] J. Gao, M. D. Sacchi, and X. Chen, "A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions," *Geophysics*, vol. 78, no. 1, pp. V21–V30, 2013.

[13] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6576–6601, 2014.

[14] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Comm. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2014.

[15] J. M. M. Anderson, G. B. Giannakis, and A. Swami, "Harmonic retrieval using higher order statistics: a deterministic formulation," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 1880–1889, Aug 1995.

[16] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Signal Process.*, vol. 38, no. 5, pp. 814–824, 1990.

[17] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.

[18] J. C. Ye, J. M. Kim, K. H. Jin, and K. Lee, "Compressive sampling using annihilating filter-based low-rank interpolation," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 777–801, 2017.

[19] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements," *Sampl. Theory Signal Image Process.*, vol. 14, no. 1, pp. 17–47, 2015.

[20] J. Cadzow, "Signal enhancement - a composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 1, pp. 49–62, 1988.

[21] J.-F. Cai, T. Wang, and K. Wei, "Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion," *Applied and Computational Harmonic Analysis*, vol. 46, no. 1, pp. 94 – 121, 2019.

[22] ——, "Spectral compressed sensing via projected gradient descent," *SIAM J. Optim*, vol. 28, no. 3, pp. 2625–2653, 2018.

[23] Y. Chen and Y. Chi, "Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 14–31, 2018.

[24] M. Fornasier, H. Rauhut, and R. Ward, "Low-rank matrix recovery via iteratively reweighted least squares minimization," *SIAM J. Optim.*, vol. 21, no. 4, pp. 1614–1640, 2011.

[25] C. Kümmerle and J. Sigl, "Harmonic mean iteratively reweighted least squares for low-rank matrix recovery," *Journal of Machine Learning Research*, vol. 19, no. 47, pp. 1–49, 2018.

[26] K. Mohan and M. Fazel, "Iterative reweighted algorithms for matrix rank minimization," *J. Mach. Learn. Res.*, vol. 13, pp. 3441–3473, 2012.

[27] M. Fazel, H. A. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," in *Proceedings of the American Control Conference*, no. 3, 2003, pp. 2156–2162.

[28] ——, "Rank minimization and applications in system theory," in *Proceedings of the American Control Conference*, no. 4, 2004, pp. 3273–3278.

[29] S. Paternain, A. Mokhtari, and A. Ribeiro, "A newton-based method for nonconvex optimization with fast evasion of saddle points," *SIAM Journal on Optimization*, vol. 29, no. 1, pp. 343–368, 2019.

[30] G. Ongie and M. Jacob, "A fast algorithm for convolutional structured low-rank matrix recovery," *IEEE Trans. Comput. Imaging*, vol. 3, no. 4, pp. 535 – 550, 2017.

[31] S. Voronin and I. Daubechies, *Functional Analysis, Harmonic Analysis, and Image Processing: A Collection of Papers in Honor of Björn Jawerth*. AMS, 2017, ch. An iteratively reweighted least squares algorithm for sparse regularization, pp. 391–411.

[32] M. Fornasier, S. Peter, H. Rauhut, and S. Worm, "Conjugate gradient acceleration of iteratively re-weighted least squares methods," *Comput. Optim. Appl.*, vol. 65, no. 1, pp. 205–259, 2016.

[33] B. Recht, "A simpler approach to matrix completion," *Journal of Machine Learning Research*, vol. 12, pp. 3413–3430, 2011.

[34] B. Vandereycken, "Low rank matrix completion by riemannian optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1214–1236, 2013.

[35] K. Wei, J.-F. Cai, T. F. Chan, and S. Leung, "Guarantees of riemannian optimization for low rank matrix completion," 2016. [Online]. Available: https://arxiv.org/abs/1603.06610

[36] B. N. Bhaskar, G. Tang, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[37] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649 – 4662, 2016.

[38] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, 2013.

[39] W. Xu, J. Yi, S. Dasgupta, J. F. Cai, M. Jacob, and M. Cho, "Separation-free super-resolution from compressed measurements is possible: an orthonormal atomic norm minimization approach," *IEEE International Symposium on Information Theory (ISIT)*, vol. 62, no. 6, pp. 3685–3701, 2018.

[40] B. B. Haro and M. Vetterli, "Sampling continuous-time sparse signals: A frequency-domain perspective," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 777–801, 2018.

[41] A. Fannjiang, "Compressive spectral estimation with single-snapshot esprit: Stability and resolution," 2016. [Online]. Available: https://arxiv.org/abs/1607.01827

[42] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.