On the convex geometry of blind deconvolution and matrix completion[*]

Felix Krahmer[†], Dominik Stöger[†]

February 28, 2019

Low-rank matrix recovery from structured measurements has been a topic of intense study in the last decade and many important problems like matrix completion and blind deconvolution have been formulated in this framework. An important benchmark method to solve these problems is to minimize the nuclear norm, a convex proxy for the rank. A common approach to establish recovery guarantees for this convex program relies on the construction of a so-called approximate dual certificate. However, this approach provides only limited insight in various respects. Most prominently, the noise bounds exhibit seemingly suboptimal dimension factors. In this paper we take a novel, more geometric viewpoint to analyze both the matrix completion and the blind deconvolution scenario. We find that for both these applications the dimension factors in the noise bounds are not an artifact of the proof, but the problems are intrinsically badly conditioned. We show, however, that bad conditioning only arises for very small noise levels: Under mild assumptions that include many realistic noise levels we derive near-optimal error estimates for blind deconvolution under adversarial noise.

1. Introduction

A number of recent works have explored the observation that various ill-posed inverse problems in signal processing, imaging, and machine learning can be naturally formulated as the task of recovering a low-rank matrix $X_0 \in \mathbb{C}^{n_1 \times n_2}$ from an underdetermined system of structured linear measurements

$$y = \mathcal{A}(X_0) + e \in \mathbb{C}^m,$$

---

[*]The results of this paper have been presented in part at the 52nd Annual Asilomar Conference on Signals, Systems, and Computers, October 28-31, 2018, Pacific Grove, USA [1]

[†]Department of Mathematics, Technische Universität München, 85748 Garching/Munich, Germany

where $\mathcal{A} : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^m$ is a linear map and $e \in \mathbb{C}^m$, $\|e\| \leq \tau$, represents additive noise. Such problems include, for example, matrix completion [2], phase retrieval [3], blind deconvolution [4], robust PCA [5], and demixing [6]. In this paper, we aim to analyze the worst case scenario, that is, we do not make any assumptions on the noise except for the bound on its Euclidean norm (this scenario is sometimes referred to as *adversarial noise*, as it allows for noise specifically designed to be most harmful in a given situation). A natural first approach to recover X_0 that remains an important benchmark is to solve the semidefinite program

$$\begin{aligned} \text{minimize} \quad & \|X\|_* \\ \text{subject to} \quad & \|y - \mathcal{A}(X)\| \leq \tau, \end{aligned}$$

where $\| \cdot \|_*$ denotes the nuclear norm, i.e., the sum of the singular values. Recovery guarantees have been shown under the assumption that the measurement operator $\mathcal{A}$ possesses a certain degree of randomness. To establish such guarantees various proof strategies have been proposed, including approaches via the restricted isometry property [7, 8], descent cone analysis [9], and so-called approximate dual certificates [10, 11]. While the latter approach remains state of the art for many structured problems including the highly relevant problems of randomized blind deconvolution and matrix completion, it seemingly has some disadvantages. Most prominently, the resulting recovery guarantees take the form

$$\|\hat{X} - X_0\|_F \lesssim \sqrt{n_1}\tau, \tag{1}$$

where $\hat{X}$ denotes a minimizer of the semidefinite program above and $\| \cdot \|_F$ denotes the Frobenius norm, whereas under comparable normalization, the first two approaches, when applicable, give rise to superior recovery guarantees of the form

$$\|\hat{X} - X_0\|_F \lesssim \tau.$$

Before this paper it was open whether the additional dimension scaling factor in (1) is a proof artifact. Similarly, for randomized blind deconvolution one of the coherence terms appearing in the result was believed to arise only from the proof technique (cf. [12, Remark 2]).

Another drawback of proceeding via an approximate dual certificate is that it gives only limited insight into geometric properties of the problems such as the null-space property [13], which is also an important ingredient for the study of some more efficient non-convex algorithms [14, 15].

Approaches via descent cone analysis [9], in contrast, provide much more geometrical insight. The underlying idea of such approaches is to study the minimum conic singular value defined by

$$\lambda_{\min}(\mathcal{A}, \mathcal{K}) := \inf_{Z \in \mathcal{K} \setminus \{0\}} \frac{\|\mathcal{A}(Z)\|}{\|Z\|_F}$$

for $\mathcal{K}$ the descent cone of the underlying atomic norm – the nuclear norm in case of low-rank matrix recovery. For a more detailed review of this approach including a precise

definition of the descent cone we refer to Section 2.3 below. Through the study of the minimum conic singular value many superior results were obtained for low-rank recovery problems, most importantly in the context of phase retrieval [16, 17]. Furthermore, minimum conic singular values can also help to understand certain nonlinear measurement models [18].

For all these reasons, it would be desirable to apply this approach also for matrix completion and blind deconvolution. A challenge that one faces, however, is that for both problems one cannot hope to recover all low-rank matrices; rather, only matrices that satisfy certain coherence constraints are admissible (cf. the discussion in [19, Section 5.4]). In this article we address this challenge, providing the first geometric analysis of these problems. We find that the dimensional factors appearing in the error bounds are the true scaling of the minimum conic singular value and hence intrinsically relate to the underlying geometry. Nevertheless for blind deconvolution, near-optimal recovery is possible, if the noise level is not too small.

1.1. Organization of the paper and our contribution

In Section 2 we will review blind deconvolution, matrix completion, as well as some techniques related to descent cone analysis. In Section 3 we will present the main results of this paper. Theorems 3.1 and 3.5 establish that for both blind deconvolution and matrix completion, nuclear norm minimization is intrinsically ill-conditioned. In contrast, Theorem 3.7 provides a near-optimal error bound for blind deconvolution when the noise level is not too small, implying that the conditioning problems only take effect for very small noise levels. The upper bounds for the minimum conic singular value which are the main ingredients of Theorems 3.1 and 3.5 are derived in Section 4. In Section 5 we prove the stability results for blind deconvolution.
We believe that not only our results, but also the proof techniques and geometric insights in this manuscript will be of general interest and help to obtain further understanding of low-rank matrix recovery models, in particular under coherence constraints. We discuss interesting directions for future research in Section 6.

2. Background and related work

2.1. Blind deconvolution

Blind deconvolution problems arise in a number of different areas in science and engineering such as astronomy, imaging, and communications. The goal is to recover both an unknown signal and an unknown kernel from their convolution. In this paper we work with the circular convolution, which is defined by

$$w * x := \left(\sum_{j=1}^{L} w_j x_{k-j} \right)_{k=1}^{L},$$

where the index difference $k - j$ is considered modulo L. Without further assumptions on w and x this bilinear map is far from injective. Consequently, it is crucial to impose structural constraints on both w and x. Arguably, the simplest such model is given by linear constraints, that is, both w and x are constrained to known subspaces. Such a model is reasonable in many applications. In wireless communication, for example, it makes sense to assume that the channel behaviour is dominated by the most direct paths and for the signal x a subspace model can be enforced by embedding the message via a suitable coding map into a higher-dimensional space before transmission.

The first rigorous recovery guarantees for such a model were derived by Ahmed, Recht, and Romberg [4]. More precisely, they assume that $w = Bh$, where $B \in \mathbb{C}^{L \times K}$ is a fixed, deterministic matrix such that $B^*B = Id_K$ (i.e., B is an isometry) and they model $x = C\overline{m_0}$, where $\overline{m_0}$ denotes the complex-conjugate of m_0. Here, the matrix $C \in \mathbb{C}^{L \times K}$ is a random matrix, whose entries are independent and identically distributed with circular symmetric normal distribution $\mathcal{CN}\left(0, \frac{1}{\sqrt{L}}\right)$. In this paper we also adopt this model.

Using the well-known fact that the Fourier transform diagonalizes the circular convolution one can rewrite

$$w * x = \sqrt{L}F^*\mathrm{diag}\left(Fw\right)Fx,$$

where $F \in \mathbb{C}^{L \times L}$ denotes the normalized, unitary discrete Fourier matrix, and

$$\widehat{w * x} := F\left(w * x\right) = \sqrt{L}\mathrm{diag}\left(FBh_0\right)FC\overline{m_0}.$$

Denoting by b_ℓ the ℓth row of the matrix $\overline{FB}$, and by c_ℓ the ℓth row of the matrix $\sqrt{L}FC$, one observes that

$$\left(\widehat{w * x}\right)_\ell = b_\ell^* h_0 m_0^* c_\ell = \mathrm{Tr}\left(h_0 m_0^* c_\ell b_\ell^*\right) = \langle b_\ell c_\ell^*, h_0 m_0^*\rangle_F.$$

Furthermore, because of the rotation invariance of the circular symmetric normal distribution all the entries of the vectors $\{c_\ell\}_{\ell=1}^L$ are (jointly) independent and identically distributed with distribution $\mathcal{CN}\left(0, 1\right)$. Noting that the expression $\langle h_0 m_0^*, b_\ell c_\ell^*\rangle_F$ is linear in $h_0 m_0^*$, Ahmed, Recht, and Romberg [4] defined the operator $\mathcal{A} : \mathbb{C}^{K \times N} \to \mathbb{C}^L$ by

$$\left(\mathcal{A}\left(X\right)\right)\left(\ell\right) := \langle b_\ell c_\ell^*, X\rangle_F \tag{2}$$

obtaining the measurement model

$$y = \widehat{w * x} + e = \mathcal{A}\left(X_0\right) + e,$$

where $e \in \mathbb{C}^L$ is additive noise and $X_0 = h_0 m_0^*$. The goal is then to determine h_0 and m_0 from $y \in \mathbb{C}^L$ up to the inherent scaling ambiguity, or, equivalently, to find the rank-one matrix $X_0 = h_0 m_0^*$.

For $e = 0$, among all solutions of the equation $y = \mathcal{A}\left(X_0\right)$, the matrix X_0 is the one with the smallest rank. For this reason, Ahmed, Recht, and Romberg [4] suggested to

minimize a natural proxy for the rank, the nuclear norm $\|\cdot\|_*$, defined as the sum of the singular values of a matrix.

$$\begin{aligned} \text{minimize} \quad & \|X\|_* \\ \text{subject to} \quad & \|\mathcal{A}(X) - y\| \leq \tau. \end{aligned} \tag{3}$$

Here $\tau > 0$ is an a priori bound for the noise level, that is, we assume that $\|e\| \leq \tau$. For this semidefinite program, they establish the following recovery guarantee.

Theorem 2.1 ([4])**.** *Consider measurements of the form $y = \mathcal{A}(h_0 m_0^*) + e$ for $h_0 \in \mathbb{C}^K$, $m_0 \in \mathbb{C}^N$, $e \in \mathbb{C}^L$, and $\mathcal{A}$ as defined in (2). Assume that $\|e\| \leq \tau$ and*

$$L/\log^3 L \gtrsim K\mu_{\max}^2 + N \max\left\{\mu_{h_0}^2; \tilde{\mu}_{h_0}^2\right\}.$$

Then with probability exceeding $1 - \mathcal{O}(L^{-1})$ every minimizer $\hat{X}$ of the SDP (3) satisfies

$$\|\hat{X} - h_0 m_0^*\|_F \lesssim \sqrt{K + N}\tau. \tag{4}$$

Here $\mu_{\max}^2$ and $\mu_{h_0}^2$ are coherence parameters, which are defined via

$$\mu_{\max}^2 := \frac{L}{K}\max_{\ell \in [L]} \|b_\ell\|^2.$$

and

$$\mu_{h_0}^2 := \frac{L}{\|h_0\|^2}\max_{\ell \in [L]} |\langle b_\ell, h_0\rangle|^2.$$

The third coherence factor $\tilde{\mu}_{h_0}$ is a technical term corresponding to a partition that is constructed as a part of the proof of Theorem 2.1, which is based on the Golfing Scheme [11].

To put the impact of the coherence factors into perspective, observe that if all vectors b_ℓ have the same ℓ_2-norm, one obtains that $\mu_{\max} = 1$; this will be the case, for example, when B is a low-frequency Fourier matrix, as it appears for applications in wireless communication. The second coherence factor always satisfies $1 \leq \mu_{h_0}^2 \leq K\mu_{\max}^2$. If μ_{h_0} is smaller, this indicates that the mass $\|h_0\|^2 = \sum_{\ell=1}^L |\langle b_\ell, h_0\rangle|^2$ is distributed fairly evenly among $|\langle b_\ell, h_0\rangle|$. For example, if $\mu_{h_0} = 1$, then $|\langle b_\ell, h_0\rangle| = \frac{1}{\sqrt{L}}\|h_0\|$ for all $\ell \in [L]$. Numerical simulations in [4] confirm that many h_0 corresponding to large μ_{h_0} show worse performance, indicating that this factor may be necessary.

The last coherence factor $\tilde{\mu}_{h_0}$, in contrast, will no longer appear in our result below, which is why we refrain from detailed discussion. We refer the interested reader to [12, Remark 2.1] and [20, Section 2.3] for details.

For generic h_0 the parameters μ_{h_0} and $\tilde{\mu}_{h_0}$ are reasonably small. For example, if h_0 is chosen from the uniform distribution on the sphere, one can show that with high probability $\mu_{h_0} = \mathcal{O}\left(\sqrt{\log L}\right)$.

For the noiseless case, i.e., $\tau = 0$, Theorem 2.1 yields exact recovery, and the required

sample complexity $L/\log^3 L \gtrsim K + N$ is optimal up to logarithmic factors, as the number of degrees of freedom is $K + N - 1$ (see [21] for an exact identifiability analysis based on algebraic geometry.) However, if there is noise, the bound for the reconstruction error scales with $\sqrt{K + N}$, in contrast to other measurement scenarios such as low-rank matrix recovery from Gaussian measurements (see, e.g., [9]).

Let us comment on some related work. The foundational paper [4] has triggered a number of follow-up works on the problem of randomized blind deconvolution. A first line of works extended the result to recovering signals from their superposition $\sum_{i=1}^{r} w_i * x_i$, a problem often referred to as blind demixing [12, 20]. Another line of works investigated non-convex (gradient-descent based) algorithms [22, 23, 24], which have the advantage that they are computationally less expensive, as they operate in the natural parameter space. It has been shown that they require a near-optimal number of measurements for recovery. For such an algorithm, [22] derived near-optimal noise-bounds for a Gaussian noise model. However, as in this paper, we focus on the scenario of adversarial noise (instead of random noise) the resulting guarantees are not comparable to ours below.

2.2. Matrix completion

The matrix completion problem of reconstructing a low-rank matrix $X_0 \in \mathbb{R}^{n_1 \times n_2}$ (we assume that w.l.o.g. $n_1 \geq n_2$) from only a part of its entries arises in many different applications such as in collaborative filtering [25] and multiclass learning [26]. For this reason one could observe a flurry of work on this problem in the last decade, and we will only be able to give a very selective overview of this topic. The precise sampling model that we consider is that m entries of X_0 are sampled uniformly at random with replacement. Denoting by e_i the standard coordinate vectors in $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, respectively, the corresponding measurement operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ can be written as

$$\mathcal{A}(X)(i) := \sqrt{\frac{n_1 n_2}{m}} \langle X, e_{a_i} e_{b_i}^* \rangle_F, \tag{5}$$

where $(a_i, b_i) \in [n_1] \times [n_2]$ is chosen uniformly at random for each $i \in [m]$ (and independently from all other measurements). The scaling factor $\sqrt{\frac{n_1 n_2}{m}}$ in the definition of the measurement operator $\mathcal{A}$ is chosen to ensure that $\mathbb{E}\left[\|\mathcal{A}(X)\|^2\right] = \|X\|_F^2$. (Some other papers on matrix completion choose a different scaling. We have chosen this normalization because in this way the results for the matrix completion problem can be better compared to those for the blind deconvolution scenario.) Alternative sampling models analyzed in other works include sampling a subset Ω uniformly from $[n_1] \times [n_2]$ (i.e., without replacement, see, e.g., [27]), or sampling using random selectors.

Again we aim to recover X_0 from noisy observations $y = \mathcal{A}(X_0) + e$, with a noise vector $e \in \mathbb{R}^m$ that satisfies $\|e\| \leq \tau$ via the SDP

$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & \|\mathcal{A}(X) - y\| \leq \tau. \end{array} \tag{6}$$

6

For matrix completion, this approach has first been studied in [2].

It is well known that similarly to the blind deconvolution problem, some incoherence assumptions are necessary to allow for successful recovery. Indeed, suppose that $X_0 = e_1 e_1^*$. Then, if $m \ll n_1 n_2$ with high probability it holds that $\mathcal{A}(X_0) = 0$ and one cannot hope to recover X_0. To avoid such special cases, one needs to ensure that the mass of the Frobenius norm of X_0 is spread out over all entries rather evenly. This property is captured by the coherence parameter [11]

$$\mu^2(X_0) := \max\left\{\mu_1^2(X_0), \mu_2^2(U), \mu_2^2(V)\right\}.$$

Here for $U \in \mathbb{R}^{n_1 \times r}$ and $V \in \mathbb{R}^{n_2 \times r}$ arising from the singular value decomposition $X_0 = U\Sigma V^*$ the three terms on the right hand side are given by

$$\mu_1(X_0) := \sqrt{\frac{n_1 n_2}{r}} \|UV^*\|_{\ell_\infty}$$

$$\mu_2(U) := \sqrt{\frac{n_1}{r}} \max_{i \in [n_1]} \|U^* e_i\|$$

$$\mu_2(V) := \sqrt{\frac{n_2}{r}} \max_{i \in [n_2]} \|V^* e_i\|.$$

For this coherence parameter, a series of works [2, 27, 11, 28] established the following recovery guarantee for the noiseless scenario.

Theorem 2.2 ([28]). *Consider measurements of the form $y = \mathcal{A}(X_0)$, where $X_0 \in \mathbb{R}^{n_1 \times n_2}$ is a rank-r matrix and $\mathcal{A}$ is given by (5). Assume that*

$$m \geq C\mu^2(X_0) r n_1 \log^2 n_1$$

Then with probability at least $1 - \mathcal{O}\left(n_1^{-1}\right)$ the matrix X_0 is the unique minimizer of the SDP (6) with $\tau = 0$.

As for blind deconvolution, this result has been shown using an approximate dual certificate. In [29] this result has been generalized to the case of adversarial noise, showing that with high probability the minimizer $\hat{X}$ of (6) satisfies

$$\|\hat{X} - X_0\|_F \lesssim \tau \sqrt{n_2}, \tag{7}$$

whenever $m \gtrsim n_1 \text{polylog } n_1$. As in the blind deconvolution framework, this error bound differs from the case of full Gaussian measurements as discussed, for example, in [9], and also from oracle estimates [5, Section III.B] by a dimensional scaling factor, which will be addressed in this paper.

Also random noise models for matrix completion have been studied in a number of works. In particular, we would like to mention [30, 31], which derive near-optimal rates (both in sample size and estimation error) for matrix completion under subexponential noise with a slightly different nuclear-norm penalized estimator than the one we consider

as long as the noise-level is not too small. Similar bounds have also been obtained in [32] using an estimator, which is closer to the one in this work.

Apart from convex methods also many nonconvex algorithms have been proposed and analysed, for example a number of variants of gradient descent (see, e.g., [33, 34, 35, 36, 37, 14, 15, 23]). Arguably the strongest result for matrix completion under adversarial noise has been shown in [33, 38]. These works propose a non-convex algorithm based on Riemannian optimization and show that if the number of measurements is larger than $r^2 n_1 \text{polylog}(n_1)$ the true matrix can be reconstructed up to an estimation error superior to the one in [29]. Namely for κ denoting the condition number of the matrix X_0 they show that the output $\hat{X}$ of their algorithm satisfies (in our notation)

$$\|\hat{X} - X_0\| \lesssim \kappa^2 \sqrt{rm} \|e\|_\infty, \tag{8}$$

provided the noise level is below a certain, small threshold that scales with the smallest singular value of X_0. For error vectors e that are spread out evenly and matrices that are well conditioned, one has that $\sqrt{m}\|e\|_\infty \approx \|e\|_2$, so this bound is superior to (7) in the sense that the scaling factors that appear only scale with the rank r and not the dimension. It should be noted though that in contrast to nuclear norm minimization the underlying algorithm requires precise knowledge of the true rank of the matrix to be recovered.

Just before completion of this manuscript, Chen et al. [39] bridged convex and nonconvex approaches, using nonconvex methods to analyze a convex recovery scheme. Their results provide near optimal recovery guarantees for the matrix completion problem via nuclear norm minimization under a subgaussian random noise model for a much larger range of admissible noise levels than the aforementioned works. More precisely, the proof is based on the observation that in their scenario the minimizer of the convex problem is very close to an approximate critical point of a non-convex gradient based method. This allows them to transfer existing stability results [23] for non-convex optimization to the convex problem. However, the required sample complexity scales suboptimally in the rank r of the matrix and similarly to (8), the error bound depends on the condition number κ.

2.3. Descent cone analysis

In recent years a number of works have studied low-rank matrix recovery and compressed sensing via a descent cone analysis. This approach has been pioneered for ℓ_1-norm minimization in [40] and for more general (atomic) norms in [9]. Here the descent cone of a norm at a point $X_0 \in \mathbb{C}^{K \times N}$ is the set of all possible directions $Z \in \mathbb{C}^{K \times N}$ such that the norm does not increase. For the nuclear norm, this leads to the following definition.

Definition 2.3. *For any matrix $X_0 \in \mathbb{C}^{K \times N}$ define its descent cone $\mathcal{K}_*(X_0)$ by*

$$\mathcal{K}_*(X_0) := \left\{ Z \in \mathbb{C}^{K \times N} : \ \|X_0 + \varepsilon Z\|_* \leq \|X_0\|_* \ for \ some \ \varepsilon > 0 \right\}.$$

To understand its relevance for recovery guarantees assume for a moment that we are in the noiseless scenario, i.e., $\tau = 0$ and $e = 0$. Then the matrix $X_0 \in \mathbb{C}^{K \times N}$ is the unique minimizer the semidefinite program (3), if and only if the null space of $\mathcal{A}$ does not intersect the descent cone $\mathcal{K}_*(X_0)$. In the case of noise, the constraint $\|y - \mathcal{A}(X_0)\| \leq \tau$ in the SDPs (3) and (6) defines a region around $X_0 + \ker \mathcal{A}$, i.e., the affine subspace consistent with the observed measurements in the noiseless scenario. The intersection of this region with the set of all signals that have a smaller nuclear norm than the ground truth X_0 is the set of feasible solutions that are preferred to X_0. The following quantity for a matrix X_0, which is often referred to as minimum conic singular value, quantifies the size of this intersection

$$\lambda_{\min}(\mathcal{A}, \mathcal{K}_*(X_0)) := \inf_{Z \in \mathcal{K}_*(X_0) \setminus \{0\}} \frac{\|\mathcal{A}(Z)\|}{\|Z\|_F}.$$

If $\lambda_{\min}(\mathcal{A}, \mathcal{K}_*(X_0))$ becomes larger, this intersection becomes smaller, which translates into stronger recovery guarantees. The following theorem confirms this intuition.

Theorem 2.4. *[9, Proposition 2.2] Let $\mathcal{A} : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^m$ be a linear operator and assume that $y = \mathcal{A}(X_0) + e$ with $\|e\| \leq \tau$. Then any minimizer $\hat{X}$ of the SDP (3) satisfies*

$$\|\hat{X} - X_0\|_F \leq \frac{2\tau}{\lambda_{\min}(\mathcal{A}, \mathcal{K}_*(X_0))}.$$

When measurement matrices of the operator $\mathcal{A}$ are full Gaussian matrices (in contrast to rank-1 measurements as in this paper) and $\mathcal{A}$ is normalized such that $\mathbb{E}[\mathcal{A}^*\mathcal{A}] = \mathrm{Id}$, for an arbitrary low-rank matrix X_0 one has with high probability that $\lambda_{\min}(\mathcal{A}, \mathcal{K}_*(X_0)) \asymp 1$. Consequently, Theorem 2.4 yields an optimal estimation error even for adversarial noise. As we will show this is no longer the case for blind deconvolution and matrix completion.

The geometric analysis of linear inverse problems via the descent cone and the minimum conic singular value has lead to many new results and insights in compressed sensing and low-rank matrix recovery. For convex programs the phase transition of the success rate could be precisely predicted [41]. As the proofs are specific to full Gaussian measuements, they do not apply for a number of important structured and heavy-tailed measurement scenarios. Stronger results [42, 43, 17, 44, 16] were subsequently obtained using Mendelson's small ball method [45, 46], a powerful tool for bounding a nonnegative empirical process from below, now often refereed to as Mendelson's small ball method.

2.4. Notation

For $n \in \mathbb{N}$ we will write $[n]$ to denote the set $\{1; \ldots; n\}$. For any set A we will denote its cardinality by $|A|$. For a complex number z we will denote its real part by $\mathrm{Re}(z)$ and its imaginary part by $\mathrm{Im}(z)$. By $\log(\cdot)$ we will denote the logarithm to the base e. By $\mathbb{E}X$ we will denote the expectation of a random variable X and by $\mathbb{P}(A)$ we denote the probability of an event A. If $v \in \mathbb{C}^n$ we will denote its ℓ_2-norm by $\|v\|$ and

its Hermitian transpose by v^*. For $u, v \in \mathbb{C}^n$ the (Euclidean) inner product is defined by $\langle u, v \rangle := u^* v$. Furthermore, for $Z \in \mathbb{C}^{n_1 \times n_2}$ its spectral norm is given by $\|Z\|$, i.e., the dual norm of the nuclear norm $\|Z\|_*$. Moreover, the Frobenius norm of Z is defined by $\|Z\|_F$ with corresponding inner product $\langle Z, W \rangle_F := \text{Tr}(Z^* W)$, where $W \in \mathbb{C}^{n_1 \times n_2}$. When we study matrix completion, we will work with matrices $Z \in \mathbb{R}^{n_1 \times n_2}$ and the previous quantities will be defined analogously. Moreover, in that scenario we will use the notation $\|Z\|_{\ell_\infty} := \max\limits_{(i,j) \in [n_1] \times [n_2]} |Z_{i,j}|$, where $\{Z_{i,j}\}_{i,j=1}^{n_1, n_2}$.

3. Our results

3.1. Instability of low-rank matrix recovery

3.1.1. Blind deconvolution

Our first main result states that randomized blind deconvolution can be unstable under adversarial noise.

Theorem 3.1. *Let $K, N \in \mathbb{N} \setminus \{1\}$. Assume that L is an integer multiple of K and that*

$$C_1 K \leq L \leq \frac{KN}{9}.$$

Then there exists a matrix $B \in \mathbb{C}^{L \times K}$ satisfying $B^* B = Id_K$ and with FB having rows of equal norm, i.e., $\mu_{\max}^2 = 1$, such that for all $h_0 \in \mathbb{C}^K \setminus \{0\}$ and $m_0 \in \mathbb{C}^N \setminus \{0\}$ the following holds:

With probability at least $1 - \mathcal{O}\left(\exp\left(-\frac{K}{C_2 \mu^2}\right)\right)$, where $\mu^2 = \mu_{h_0}^2 = \frac{L}{\|h_0\|^2} \max\limits_{\ell \in [L]} |\langle b_\ell, h_0 \rangle|^2$, there is $\tau_0 > 0$ such that for all $\tau \leq \tau_0$ there exists an adversarial noise vector $e \in \mathbb{C}^L$ with $\|e\| \leq \tau$ that admits an alternative solution $\tilde{X}$ with the following properties.

- *$\tilde{X}$ is feasible, i.e., $\|\mathcal{A}\left(\tilde{X}\right) - y\| = \tau$ for $y = \mathcal{A}\left(h_0 m_0^*\right) + e$ the noisy measurement vector*

- *$\tilde{X}$ is preferred to $X_0 = h_0 m_0^*$ by the SDP (3), i.e., $\|\tilde{X}\|_* \leq \|X_0\|_*$, but*

- *$\tilde{X}$ is far from the true solution in Frobenius norm, i.e.,*

$$\|\tilde{X} - X_0\|_F \geq \frac{\tau}{C_3} \sqrt{\frac{KN}{L}}.$$

The constants C_1, C_2, and C_3 are universal.

Remark 3.2. *The matrix B in the above result exactly fits into the framework of Theorem 2.1, and also one can check that for our choice of B (see the proof of Proposition 3.3 for its definition) one has that $\mu_{2,h_0} = \mu_{1,h_0}$. That is, the assumptions of Theorem 2.1 cannot be enough to deduce stability.*

We do not expect, however, that this kind of instability is observed for arbitrary isometric embeddings $B \in \mathbb{C}^{L \times K}$. In particular, if B is a random embedding we expect that a similar result as in [17] applies.

To put our results in perspective note that for $L \asymp (K + N)\, \text{polylog}\,(K + N)$, which is the minimal number of measurements required for noiseless recovery, it holds that $\sqrt{\frac{KN}{L}} \asymp \sqrt{\frac{\min\{K,N\}}{\text{polylog}(K+N)}}$. Up to logarithmic factors, this coincides with the rate predicted by (4), whenever $K \asymp N$.

Theorem 3.1 is a direct consequence of the following proposition, which we think is interesting in its own right.

Proposition 3.3. *Let $K, N \in \mathbb{N} \setminus \{1\}$. Assume that L is an integer multiple of K and that*

$$C_1 K \leq L \leq \frac{KN}{9}. \tag{9}$$

Then there exists $B \in \mathbb{C}^{L \times K}$ satisfying $B^* B = Id_K$ and $\mu_{\max}^2 = 1$, whose corresponding measurement operator $\mathcal{A}$ satisfies the following.

Let $h_0 \in \mathbb{C}^K \setminus \{0\}$, $m_0 \in \mathbb{C}^N \setminus \{0\}$ and set $\mu^2 := \mu_{h_0} = L \frac{\max_{\ell \in [L]} |\langle b_\ell, h_0 \rangle|^2}{\|h_0\|^2}$. Then with probability at least $1 - \mathcal{O}\left(\exp\left(-C_2 K/\mu^2\right)\right)$ it holds that

$$\lambda_{\min}\left(\mathcal{A}, \mathcal{K}_*\left(h_0 m_0^*\right)\right) \leq C_3 \sqrt{\frac{L}{KN}}. \tag{10}$$

Here C_1, C_2, and C_3 are absolute constants.

The proof of Proposition 3.3 will be provided in Section 4. Note that by definition of the minimum conic singular value $\lambda_{\min}\left(\mathcal{A}, \mathcal{K}_*\left(h_0 m_0^*\right)\right)$ Proposition 3.3 is equivalent to the statement that with high probability there is $Z \in \mathcal{K}_*\left(h_0 m_0^*\right) \setminus \{0\}$ such that

$$\|\mathcal{A}\left(Z\right)\| \lesssim \sqrt{\frac{L}{KN}} \|Z\|_F.$$

Our construction of such $Z \in \mathcal{K}_*\left(h_0 m_0^*\right)$ relies on the observation that with high probability there is a rank-one matrix $W \in \mathbb{C}^{K \times N}$ in the null-space of $\mathcal{A}$ which is relatively close to the descent cone (with respect to the $\|\cdot\|_F$-distance). Perturbing W by $-\beta h_0 m_0^*$ for a suitable β one can then obtain a matrix $Z \in \mathcal{K}_*\left(h_0 m_0^*\right)$, which fulfills (3.1.1).
The existence of such a matrix $W \in \ker \mathcal{A}$ also reveals a fact about the geometry of the problem, which we find somewhat surprising: while the null space of $\mathcal{A}$ does not intersect the descent cone (otherwise exact recovery would not be possible), the angle between those objects is very small. This is very different from the behavior for measurement matrices $\mathcal{A}$ with i.i.d. Gaussian entries (instead of $b_\ell c_\ell^*$).

Remark 3.4. *While $\tilde{X}$ is preferred to the true solution by the SDP (3) $\tilde{X}$ is typically not a minimizer of (3). To see this, assume that without noise exact recovery is possible, which is the case with high probability by Theorem 2.1. Then consider $\tilde{X} = X_0 + tZ$ for $Z \in \mathcal{K}_*\left(h_0 m_0^*\right)$ of the form $Z = W - \beta h_0 m_0^*$ with $W \in \ker \mathcal{A}$ and $\beta > 0$ such that*

$\frac{\|\mathcal{A}(Z)\|}{\|Z\|_F} \lesssim \sqrt{\frac{L}{KN}}$, as in the proof of Proposition 3.3. As $W \notin \mathcal{K}_* (h_0 m_0^*)$ (otherwise exact recovery would not be possible) it follows that for $t > 0$

$$\begin{aligned} \|\tilde{X}\|_* &= \|X_0 + tZ\|_* \\ &= \|(1 - t\beta) X_0 + tW\|_* \\ &> \|(1 - t\beta) X_0\|_* \end{aligned}$$

where the last line is due to $\mathcal{K}_* (X_0) = \mathcal{K}_* ((1 - t\beta) X_0)$.

On the other hand, we also have that $\mathcal{A}\left(\hat{X}\right) = \mathcal{A}((1 - t\beta) X_0)$ due to $\mathcal{A}(W) = 0$ and, hence, $(1 - t\beta) X_0$ is admissible whenever $\tilde{X}$ is admissible. Consequently, the SDP (3) will always prefer $(1 - \beta t) h_0 m_0^*$ to $\tilde{X}$ and $\tilde{X}$ will never be a minimizer. It remains an open problem what one can say about the minimizer $\hat{X}$ of (3), see also Section 6. Even if the minimizer of (3) $\hat{X}$ is closer to the ground truth (in $\|\cdot\|_F$-distance) than $\tilde{X}$, however, the nuclear norms of X and $\tilde{X}$ will be very close, which can easily lead to numerical instabilities.

3.1.2. Matrix completion

Our second main result states that for arbitrary incoherent low-rank matrices, matrix completion is unstable with high probability. Note that in contrast to Theorem 3.1 which is based on a specific choice of parameters the following result holds for an arbitrary incoherent matrix X_0.

Theorem 3.5. Let $n_1 \geq n_2$ and let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be defined as in (5). Assume that $X_0 \in \mathbb{R}^{n_1 \times n_2} \setminus \{0\}$ is a rank r matrix with singular value decomposition $X_0 = U\Sigma V^*$. Moreover, assume that

$$C_1 r n_1 \mu_2^2 (V) \log(2r) \leq m \leq \frac{n_1 n_2}{32}.$$

Then with probability at least $1 - \mathcal{O}\left(\exp\left(-\frac{m}{C_2 r \max\{\mu_1^2(X_0), \mu_2^2(V)\}}\right)\right)$ there is $\tau_0 > 0$ such that for all $\tau \leq \tau_0$ there exists an adversarial noise vector $e \in \mathbb{R}^m$ with $\|e\| \leq \tau$ that admits an alternative solution $\tilde{X} \in \mathbb{R}^{n_1 \times n_2}$ with the following properties.

- $\tilde{X}$ is feasible, i.e., $\left\|\mathcal{A}\left(\tilde{X}\right) - y\right\| = \tau$ for $y = \mathcal{A}(X_0) + e$ the noisy measurement vector

- $\tilde{X}$ is preferred to X_0 by the SDP (6), i.e., $\|\tilde{X}\|_* \leq \|X_0\|_*$, but

- $\tilde{X}$ is far from the true solution in Frobenius norm, i.e.,

$$\|\tilde{X} - X_0\|_F \geq \frac{\tau}{C_3} \sqrt{\frac{r n_1 n_2}{m}}.$$

Here the constants C_1, C_2, and C_3 are universal.

Again, to put our results in perspective note that for $m \asymp n_1 \text{polylog}(n_1)$, which is the minimal number of measurements required for noiseless recovery, it holds that $\sqrt{\frac{rn_1n_2}{m}} \asymp \sqrt{\frac{n_2}{\text{polylog}(n_1)}}$. Up to logarithmic factors, this coincides with the rate predicted by (7).

Theorem 3.5 is a direct consequence of the following proposition, which in our opinion is of independent interest, as it provides a negative answer to a question by Tropp [19, Section 5.4].

Proposition 3.6. *Let $X_0 \in \mathbb{R}^{n_1 \times n_2} \setminus \{0\}$ be a rank-r matrix with corresponding singular value decomposition $X_0 = U\Sigma V^*$. Moreover, assume that*

$$C_1 r n_1 \mu_2^2(V) \log(2r) \leq m \leq \frac{n_1 n_2}{32}. \tag{11}$$

Then with probability at least $1 - \mathcal{O}\left(\exp\left(-\frac{m}{C_2 \max\{\mu_1^2(X_0), r\mu_2^2(V)\}}\right)\right)$ it holds that

$$\lambda_{\min}(\mathcal{A}, \mathcal{K}_*(X_0)) \leq C_3 \sqrt{\frac{m}{n_1 n_2 r}}. \tag{12}$$

The constants C_1, C_2, and C_3 are universal.

Proposition 3.6 corresponds to Proposition 3.3 for blind deconvolution and will be proved analogously. We will again show that with high probability there is $W \in \mathbb{R}^{n_1 \times n_2}$ such that $\mathcal{A}(W) = 0$ and W is relatively close to the descent cone of X_0 in $\|\cdot\|_F$-distance. Setting $Z := W - \beta UV^*$ for a suitable $\beta > 0$ yields an element of $\mathcal{K}_*(X_0)$ with

$$\frac{\|\mathcal{A}(Z)\|}{\|Z\|_F} \leq C_3 \sqrt{\frac{m}{n_1 n_2 r}}.$$

3.2. Stable recovery

A geometric interpretation of Theorems 3.1 and 3.5 is that the nuclear norm ball is near-tangential to both the kernels of matrix completion and randomized blind deconvolution. Given that tangent spaces only provide local approximation, these results leave open, what happens in some distance, i.e., for larger noise levels – this will depend on the curvature of the nuclear norm ball.

Our third main result concerns exactly this problem for the randomized blind deconvolution setup. As it turns out, the descent directions $Z \in \mathcal{K}_*(h_0 m_0^*)$ with $\|\mathcal{A}(Z)\|/\|Z\|_F$ very small correspond to directions of significant curvature. That is, only a very short segment in this direction will have smaller nuclear norm than $h_0 m_0^*$, and the corresponding alternative solutions all correspond to very small e. For noise levels τ large enough, in contrast, these directions can be excluded and one can obtain near-optimal error bounds. In order to precisely formulate this observation, we denote the set of μ-incoherent vectors $h \in \mathbb{C}^K$ with respect to $B \in \mathbb{C}^{L \times K}$ for $\mu \geq 1$ by

$$\mathcal{H}_\mu := \left\{ h_0 \in \mathbb{C}^K : \quad \sqrt{L}|\langle b_\ell, h_0 \rangle| \leq \mu \|h_0\| \text{ for all } \ell \in [L] \right\}.$$

13

With this notation, our result reads as follows.

Theorem 3.7. *Let $\alpha > 0$ and $B \in \mathbb{C}^{L \times K}$ such that $B^*B = \text{Id}$. Assume that*

$$L \geq C_1 \frac{\mu^2}{\alpha^2} (K + N) \, \log^2 L.$$

Then with probability at least $1 - \mathcal{O}\left(\exp\left(-\frac{L\alpha^{4/3}}{C_2 \log^{4/3}(eL)\mu^{4/3}}\right)\right)$ the following statement holds for all $h_0 \in \mathcal{H}_\mu \setminus \{0\}$, all $m_0 \in \mathbb{C}^N \setminus \{0\}$, all $\tau > 0$, and all $e \in \mathbb{C}^L$ with $\|e\| \leq \tau$: Any minimizer $\hat{X}$ of (3) satisfies

$$\|\hat{X} - h_0 m_0^*\|_F \leq \frac{C_3 \mu^{2/3} \log^{2/3} L}{\alpha^{2/3}} \max\left\{\tau; \alpha\|h_0 m_0^*\|_F\right\}.$$

Here C_1, C_2, and C_3 are absolute constants.

In words, this theorem establishes linear scaling in the noise level τ with only a logarithmic dimensional factor for $\tau \geq \alpha\|h_0 m_0^*\|_F$, in contrast to the polynomial factor required for small noise levels as a consequence of Theorem 3.1. Here the value of α can be chosen arbitrarily small, at the expense of an increased number of measurements. For example when one is interested in noise levels $\tau = \epsilon\mu^{-2} \log^{-2} L$ for some $\epsilon > \epsilon_0$ (this is the largest order to expect meaningful error bounds despite the additional logarithmic factors) one should choose $\alpha \asymp \epsilon_0 \log^{-2} L$, and near-linear error bounds will be guaranteed for a sample complexity of

$$L \geq C_1 \frac{\mu^6}{\epsilon_0^2} (K + N) \, \log^6 L.$$

Remark 3.8. *A similar approach to the proof of Theorem 3.7 also yields a corresponding result for rank-one matrix completion. Arguably, however, matrix completion is mainly of interest for ground truth matrices of rank higher than one, so we decided to omit the proof details.*

4. Upper bounds for the minimum conic singular values

4.1. Characterization of the descent cone of the nuclear norm

The goal of this section is to prove Proposition 3.3 and Proposition 3.6, from which we will then be able to deduce Theorem 3.1 and Theorem 3.5. For that we first discuss a characterization of the descent cone $\mathcal{K}_*(X)$. In order to state this characterization, Lemma 4.1, we need to introduce some additional notation. Let $X \in \mathbb{C}^{n_1 \times n_2}$ be a matrix of rank r. We will denote its corresponding singular value decomposition by $X = U\Sigma V^*$, where $\Sigma \in \mathbb{R}^{r \times r}$ is a diagonal matrix with nonnegative entries and $U \in \mathbb{C}^{n_1 \times r}$ and $V \in \mathbb{C}^{n_2 \times r}$ are unitary matrices, i.e., $U^*U = V^*V = \text{Id}_r$. This allows us to define the tangent space of the manifold of rank-r matrices at the point X by

$$T_X := \left\{UA^* + BV^* : \ A \in \mathbb{C}^{n_2 \times r}, B \in \mathbb{C}^{n_1 \times r}\right\}. \tag{13}$$

By $\mathcal{P}_{T_X}$ we will denote the orthogonal projection onto T_X, by $\mathcal{P}_{T_X^\perp} = \text{Id} - \mathcal{P}_{T_X}$ the projection onto its orthogonal complement.

Lemma 4.1. [1] *Let $X \in \mathbb{C}^{n_1 \times n_2} \backslash \{0\}$ be a matrix of rank r with corresponding singular value decomposition $X = U\Sigma V^*$. Then*

$$\overline{\mathcal{K}_*(X)} = \left\{ Z \in \mathbb{C}^{n_1 \times n_2} : -Re\left(\langle UV^*, Z\rangle_F\right) \geq \|\mathcal{P}_{T_X^\perp}(Z)\|_*\right\},$$

where $\overline{\mathcal{K}_*(X)}$ denotes the topological closure of $\mathcal{K}_*(X)$.

The proof of Lemma 4.1 relies on the duality between the descent cone and the subdifferential of a convex function. In the following we will denote by $\partial \|\cdot\|_*(X)$ the subdifferential of the nuclear norm at the point $X \in \mathbb{C}^{n_1 \times n_2}$. We will use that a characterization of $\partial \|\cdot\|_*$ is well-known [47]. Namely, for all $X \in \mathbb{C}^{n_1 \times n_2}$ with corresponding singular value decomposition $X = U\Sigma V^*$ it holds that

$$\partial \|\cdot\|_*(X) = \left\{ W \in \mathbb{C}^{n_1 \times n_2} : \ \mathcal{P}_{T_X} W = UV^*, \ \|\mathcal{P}_{T_X^\perp} W\| \leq 1\right\}. \tag{14}$$

Proof. Recall that for a set of matrices $\mathcal{V} \subset \mathbb{C}^{n_1 \times n_2}$ its polar cone $\mathcal{V}^\circ$ is defined by

$$\mathcal{V}^\circ := \left\{ Z \in \mathbb{C}^{n_1 \times n_2} : \text{Re}\left(\langle W, Z\rangle_F\right) \leq 0 \text{ for all } W \in \mathcal{V}\right\}.$$

For all $X \in \mathbb{C}^{n_1 \times n_2} \backslash \{0\}$ we have the following polarity relation between the descent cone and the subdifferential (see, e.g., [48, Theorem 23.7] [2])

$$\mathcal{K}_*(X)^\circ = \overline{\{\lambda W : \ \lambda \geq 0, \ W \in \partial\|\cdot\|_*(X)\}}.$$

Thus, it follows from the bipolar theorem (see, e.g., [49, p. 53]) that

$$\overline{\mathcal{K}_*(X)} = (\partial\|\cdot\|_*(X))^\circ.$$

Hence, in order to complete the proof it is sufficient to show that

$$\left\{ Z \in \mathbb{C}^{n_1 \times n_2} : -\text{Re}\left(\langle UV^*, Z\rangle_F\right) \geq \|\mathcal{P}_{T_X^\perp}(Z)\|_*\right\} = (\partial\|\cdot\|_*(X))^\circ = \overline{\text{cone}(\partial\|\cdot\|_*(X))}. \tag{15}$$

First, suppose that $Z \in \mathbb{C}^{n_1 \times n_2}$ satisfies $-\text{Re}\left(\langle UV^*, Z\rangle_F\right) \geq \|\mathcal{P}_{T_X^\perp}(Z)\|_*$. We have to show that $\text{Re}\left(\langle W, Z\rangle_F\right) \leq 0$ for all $W \in \partial\|\cdot\|_*(X)$. Indeed,

$$\begin{aligned}
\text{Re}\left(\langle W, Z\rangle_F\right) &= \text{Re}\left(\langle \mathcal{P}_{T_X} W, Z\rangle_F\right) + \text{Re}\left(\langle \mathcal{P}_{T_X^\perp} W, Z\rangle_F\right) \\
&= \text{Re}\left(\langle UV^*, Z\rangle_F\right) + \text{Re}\left(\langle \mathcal{P}_{T_X^\perp} W, \mathcal{P}_{T_X^\perp} Z\rangle_F\right) \\
&\leq \text{Re}\left(\langle UV^*, Z\rangle_F\right) + \|\mathcal{P}_{T_X^\perp} W\| \|\mathcal{P}_{T_X^\perp} Z\|_* \\
&\leq \text{Re}\left(\langle UV^*, Z\rangle_F\right) + \|\mathcal{P}_{T_X^\perp} Z\|_* \\
&\leq 0.
\end{aligned}$$

---

[1] Lemma 4.1 is similar to well-known results in convex optimization and may be known to the community. As we could not find it in the literature in this form, we decided to include a proof for completeness.

[2] Reference [48] considers only sets and functions defined in $\mathbb{R}^n$ with the usual Euclidean inner product. However, note that $\mathbb{C}^{n_1 \times n_2}$ together with the inner product given by $\text{Re}\left(\langle \cdot, \cdot\rangle_F\right)$ may be identified with an $2n_1 n_2$-dimensional real-valued vector space with standard Euclidean inner product. Applying the results of [48] to this vector space yields the above-mentioned results.

In the first inequality we have used that the spectral norm is the dual norm of the nuclear norm. The second inequality follows from $\|\mathcal{P}_{T_X^\perp} W\| \leq 1$. Hence, we have shown that $Z \in (\partial \| \cdot \|_* (X))^\circ$. Next, let $Z \in (\partial \| \cdot \|_* (X))^\circ$ be arbitrary. Choose $\tilde{W} \in T_X^\perp$ such that $\mathrm{Re}\left(\langle \tilde{W}, Z \rangle_F \right) = \|\mathcal{P}_{T_X^\perp} (Z)\|_*$ and $\|\tilde{W}\| \leq 1$. Then by (14) it follows that $UV^* + \tilde{W} \in \partial \| \cdot \|_* (X)$ and as $Z \in (\partial \| \cdot \|_* (X))^\circ$ we obtain that

$$
\begin{aligned}
0 &\geq \mathrm{Re}\left(\langle UV^* + \tilde{W}, Z \rangle_F \right) \\
&= \mathrm{Re}\left(\langle UV^*, Z \rangle_F \right) + \|\mathcal{P}_{T_X^\perp} (Z)\|_*.
\end{aligned}
$$

This shows that $-\mathrm{Re}\left(\langle UV^*, Z \rangle_F \right) \geq \|\mathcal{P}_{T_X^\perp} (Z)\|_*$. Hence, we have verified (15), which completes the proof. $\qquad\square$

4.2. Upper bound for blind deconvolution

The goal of this section is to prove Proposition 3.3. For that we need the following lemma, which is a consequence of the concentration of measure theorem for Lipschitz functions. (For a proof of the real-valued case see, e.g., [50, Lemma 5.3.2]. The complex-case can be shown analogously.)

Lemma 4.2. *Let $P : \mathbb{C}^n \to \mathbb{C}^n$ be a random projection onto a k-dimensional subspace, which is uniformly distributed in the Grassmannian $Gr\left(k, \mathbb{C}^n\right)$. Fix $z \in \mathbb{C}^n$. Then for all $\varepsilon > 0$ with probability at least $1 - 2e^{-\tilde{c}k\varepsilon^2}$ we have that*

$$
(1 - \varepsilon) \frac{k}{n} \|z\|^2 \leq \|Pz\|^2 \leq (1 + \varepsilon) \frac{k}{n} \|z\|^2,
$$

where $\tilde{c} > 0$ is absolute some constant.

Proof of Proposition 3.3. By assumption we may write $L = \eta K$ for some $\eta \in \mathbb{N}$. Let $\{e_i\}_{i=1}^K$ an arbitrary orthonormal basis of $\mathbb{C}^K$. Choose $B \in \mathbb{C}^{L \times K}$ such that the matrix $\overline{FB}$ consists of rows $\{b_{k,i}\}_{k,i}^{K,\eta}$ defined by $b_{k,i} = \frac{1}{\sqrt{\eta}} e_k$ for all $k \in [K]$ and all $i \in [\eta]$. We observe that $(FB)^* (FB) = \mathrm{Id}_K$, which implies $B^*B = \mathrm{Id}_K$. Furthermore, we have $\|b_{k,i}\| = \sqrt{\frac{1}{\eta}}$ for all $k \in [K]$ and $i \in [\eta]$, which implies that also all rows of FB have equal norm.

Without loss of generality we assume that $\|h_0\| = \|m_0\| = 1$ as rescaling does not change the descent cone $\mathcal{K}_*(h_0 m_0^*)$. For the proof we will condition on two events. The first event states that

$$
\|\mathcal{A}(h_0 m_0^*)\| \leq 2\|h_0 m_0^*\|_F, \tag{16}
$$

which by the Bernstein inequality (see, e.g., [50]) is fulfilled with probability at least $1 - \exp\left(-cL/\mu^2\right)$, where $c > 0$ is some numerical constant. To formulate the second event, we define for all $i \in [K]$ the subspaces

$$
D_i := \mathrm{span}\{c_{i,1}; \ldots; c_{i,\eta}\} \subset \mathbb{C}^N
$$

and denote by $m_i^{\parallel}$ the orthogonal projection of m_0 onto D_i and by $m_i^{\perp}$ the projection onto $D_i^{\perp}$, the orthogonal complement of D_i. Note that $D_i \subset \mathbb{C}^N$ is a random subspace of dimension $\frac{L}{K}$, distributed uniformly over the Grassmannian $Gr\left(\frac{L}{K}, \mathbb{C}^N\right)$ due to the rotation invariance of $\mathcal{CN}(0,1)$. Hence, as $\|m_0\| = 1$ Lemma 4.2 yields that for fixed $i \in [K]$ with probability at least $1 - 2\exp\left(-\frac{\tilde{c}L}{4K}\right)$ one has

$$\frac{L}{2KN} \leq \|m_i^{\parallel}\|^2 \leq \frac{3L}{2KN}. \tag{17}$$

As the matrix C is Gaussian, the different subspaces D_i's and hence also the random vectors $\left\{m_i^{\parallel}\right\}_{i=1}^K$ are independent, so with probability at least

$$1 - \left(2\exp\left(-\frac{\tilde{c}L}{4K}\right)\right)^K = 1 - \exp\left(K\log 2 - \frac{\tilde{c}L}{4}\right)$$

there exists at least one $k \in [K]$ such that (17) holds (with $k = i$). Also note that

$$1 - \exp\left(K\log 2 - \frac{\tilde{c}L}{4}\right) \geq 1 - \exp\left(-\frac{C_1}{2}L\right),$$

which for $C_1 = \frac{8\log 2}{\tilde{c}}$ follows from assumption (9).

To summarize, we have shown that the two events $\mathcal{E}_1 := \{\|\mathcal{A}(h_0 m_0^*)\| \leq 2\|h_0 m_0^*\|_F\}$ and

$$\mathcal{E}_2 := \left\{\exists k \in [K] : \frac{L}{2KN} \leq \|m_k^{\parallel}\|^2 \leq \frac{3L}{2KN}\right\}$$

happen with probability at least $1 - \mathcal{O}\left(\exp\left(-C_2 L/\mu^2\right)\right)$, where $C_2 > 0$ is an appropriately chosen constant.

Conditional on $\mathcal{E}_1$ and $\mathcal{E}_2$, we will construct $Z \in \mathbb{C}^{K \times N}$ (depending on the vectors $(c_{i,k})_{i,k}$) such that $Z \in \overline{\mathcal{K}_*(h_0 m_0^*)} \setminus \{0\}$ and such that the inequality

$$\|\mathcal{A}(Z)\| < 8\sqrt{\frac{L}{KN}}\|Z\|_F \tag{18}$$

is satisfied. Note that this will complete the proof. Indeed, by definition of the closure and the continuity of $\mathcal{A}$ this implies that there exists $\tilde{Z} \in \mathcal{K}_*(X_0)$ such that

$$\frac{\|\mathcal{A}(\tilde{Z})\|}{\|\tilde{Z}\|_F} \leq 8\sqrt{\frac{L}{KN}},$$

which by the definition of $\lambda_{\min}(\mathcal{A}, \mathcal{K}_*(h_0 m_0^*))$ implies that (10) holds with constant $C_3 = 8$.

To construct Z satisfying (18), define

$$W := -\frac{\langle h_0, e_i \rangle}{\|m_i^\perp\| |\langle e_i, h_0 \rangle|} e_i \left(m_i^\perp \right)^*,$$

where $i \in [K]$ is chosen to satisfy (17). It follows directly from the definition of W that $\|W\|_F = 1$. We observe that $\mathcal{A}(W) = 0$ as for each $k \in [K]$ and $i \in [\eta]$ we either have $\langle e_k, b_{k,i} \rangle = 0$ or $\langle m_{k,i}^\perp, c_{k,i} \rangle = 0$. Denote by $T = T_{X_0}$ the tangent space of the manifold of rank-one matrices at $X_0 = h_0 m_0^*$ as defined in (13) and by $\mathcal{P}_T$ and $\mathcal{P}_{T^\perp}$ the corresponding orthogonal projections. It follows that

$$
\begin{aligned}
\|\mathcal{P}_{T^\perp} W\|_* &= \left\| \mathcal{P}_{T^\perp} \left(e_i \left(\frac{m_i^\perp}{\|m_i^\perp\|} \right)^* \right) \right\|_F \\
&= \left\| P_{h_0^\perp} e_i \right\| \left\| P_{m_0^\perp} \left(\frac{m_i^\perp}{\|m_i^\perp\|} \right) \right\| \\
&= \sqrt{1 - |\langle h_0, e_i \rangle|^2} \sqrt{1 - \left| \langle m_0, \frac{m_i^\perp}{\|m_i^\perp\|} \rangle \right|^2} \\
&\leq \sqrt{1 - \left| \langle m_0, \frac{m_i^\perp}{\|m_i^\perp\|} \rangle \right|^2} \\
&= \sqrt{1 - \|m_i^\perp\|^2} \\
&= \|m_i^\|\| \\
&\overset{(17)}{\leq} \sqrt{\frac{3L}{2KN}}.
\end{aligned}
\tag{19}
$$

Thus we have shown that W, an element of the null space of $\mathcal{A}$, is close to the tangent space T. We will now show that for $\beta = 2\sqrt{\frac{L}{KN}}$

$$Z := -\beta h_0 m_0^* + W,$$

lies in the closure of the descent cone $\overline{\mathcal{K}_* (h_0 m_0^*)}$. For that, we observe that

$$
\begin{aligned}
-\mathrm{Re}\left(\langle Z, h_0 m_0^* \rangle \right) &= \beta - \mathrm{Re}\left(\langle W, h_0 m_0^* \rangle_F \right) \\
&= \beta + \mathrm{Re}\left(\frac{\langle h_0, e_i \rangle}{\|m_i^\perp\| |\langle h_0, e_i \rangle|} \langle e_i \left(m_i^\perp \right)^*, h_0 m_0^* \rangle_F \right) \\
&= \beta + \mathrm{Re}\left(\frac{\langle h_0, e_i \rangle}{\|m_i^\perp\| |\langle h_0, e_i \rangle|} \langle e_i, h_0 \rangle \langle m_0, m_i^\perp \rangle \right) \\
&= \beta + |\langle h_0, e_i \rangle| \|m_i^\perp\| \\
&\geq \beta \\
&\overset{(19)}{\geq} \|\mathcal{P}_{T^\perp} W\|_* \\
&= \|\mathcal{P}_{T^\perp} Z\|_*
\end{aligned}
$$

and hence Lemma 4.1 entails that $Z \in \overline{\mathcal{K}_* (h_0 m_0^*)}$. Moreover, note that by the triangle inequality and by the assumption $L \leq \frac{1}{16} KN$ it holds that

$$
\begin{aligned}
\|Z\|_F &\geq \|W\|_F - \beta \\
&= 1 - 2\sqrt{\frac{L}{KN}} \\
&> \frac{1}{2}.
\end{aligned}
\tag{20}
$$

These observations together with $\mathcal{A}(W) = 0$ yield that

$$
\|\mathcal{A}(Z)\| = \|\mathcal{A}(\beta h_0 m_0^*)\| \overset{(16)}{\leq} 2\beta = 4\sqrt{\frac{L}{KN}} \overset{(20)}{<} 8\sqrt{\frac{L}{KN}} \|Z\|_F.
$$

This shows (18), as desired.

$\square$

4.3. Upper bound for matrix completion

In this section, we prove Proposition 3.6. For that we introduce sets $\mathcal{N}_a$, $a \in [n_1]$, via

$$
\mathcal{N}_a := \{b \in [n_2] : a = a_i \text{ and } b = b_i \text{ for some } i \in [m]\}.
$$

That is, $\mathcal{N}_a$ contains all the indices of the ath row of the matrix X_0, which are observed by the measurements. Furthermore, define by $P_{\mathcal{N}_a} \in \mathbb{R}^{n_2 \times n_2}$ the projection onto the coordinates, which are contained in $\mathcal{N}_a$, i.e. $P_{\mathcal{N}_a} = \sum_{b \in \mathcal{N}_a} e_b e_b^*$. By $P_{\mathcal{N}_a^\perp} = \sum_{b \in [m] \setminus \mathcal{N}_a} e_b e_b^*$ we denote the coordinate projection onto $[m] \setminus \mathcal{N}_a$.

We need the following technical lemma.

Lemma 4.3. *Let $V \in \mathbb{R}^{n_2 \times r}$ be an isometry, i.e. $V^* V = Id$. Assume that*

$$
m \geq C_1 r n_1 \mu_2^2(V) \log(2r).
\tag{21}
$$

Then with probability at least $1 - \mathcal{O}\left(\exp\left(-\frac{m}{C_2 r \mu_2^2(V)}\right)\right)$ there exists $a \in [n_1]$ such that

$$
\|P_{\mathcal{N}_a} V\| \leq \sqrt{\frac{2m}{n_1 n_2}}.
$$

C_1 and C_2 are universal constants.

Proof. For each $a \in [n_1]$ we set $\mathcal{I}_a := \{i \in [m] : a_i = a\}$ and define the event

$$
\mathcal{E}_a := \left\{\|P_{\mathcal{N}_a} V\|^2 \leq \frac{2m}{n_1 n_2}\right\}.
$$

We will first derive a lower bound for $\mathbb{P}\left(\mathcal{E}_a | \mathcal{I}_a\right)$. For that we note that $\|P_{\mathcal{N}_a} V\|^2 = \|V^* P_{\mathcal{N}_a} V\|$. Let $v_1, v_2, \ldots, v_{n_2}$ denote the rows of the matrix V. By definition of $\mathcal{N}_a$ and $\mathcal{I}_a$ it follows that for $X_i := v_{b_i} v_{b_i}^*$

$$V^* P_{\mathcal{N}_a} V = \sum_{b \in \mathcal{N}_a} v_b v_b^* \preceq \sum_{i \in \mathcal{I}_a} v_{b_i} v_{b_i}^* = \sum_{i \in \mathcal{I}_a} X_i. \tag{22}$$

Here we write $A \preceq B$ for two symmetric matrices A and B, if and only if $B - A$ is positive semidefinite. By (22) it is sufficient to bound the probability of the event

$$\left\{ \Big\| \sum_{i \in \mathcal{I}_a} X_i \Big\| \leq \frac{2m}{n_1 n_2} \right\} \subset \mathcal{E}_a \tag{23}$$

conditionally on $\mathcal{I}_a$. To bound $\| \sum_{i \in \mathcal{I}_a} X_i \|$ we will use the matrix Bernstein inequality (see, e.g., [51, Theorem 6.1.1]) conditionally on $\mathcal{I}_a$, which requires as ingredients

$$\mathbb{E}\left[\sum_{i \in \mathcal{I}_a} X_i \, | \mathcal{I}_a \right] = \frac{|\mathcal{I}_a|}{n_2} \sum_{b=1}^{n_2} v_b v_b^* = \frac{|\mathcal{I}_a|}{n_2} V^* V = \frac{|\mathcal{I}_a|}{n_2} \mathrm{Id},$$

an upper bound for $\sigma^2\left(\mathcal{I}_a\right) := \left\| \mathbb{E}\left[\sum_{i \in \mathcal{I}_a} \left(X_i - \mathbb{E}\left[X_i\right]\right)^2 | \mathcal{I}_a \right] \right\|$ and a constant $K > 0$ such that $\|X_i - \mathbb{E}\left[X_i\right]\| \leq K$ almost surely. To bound $\sigma^2\left(\mathcal{I}_a\right)$ we note that

$$\mathbb{E}\left[\sum_{i \in \mathcal{I}_a} \left(X_i - \mathbb{E}\left[X_i\right]\right)^2 | \mathcal{I}_a \right] \preceq \mathbb{E}\left[\sum_{i \in \mathcal{I}_a} X_i^2 | \mathcal{I}_a \right]$$

$$\preceq \left(\max_{b \in [n_2]} \|v_b\|^2 \right) \mathbb{E}\left[\sum_{i \in \mathcal{I}_a} X_i | \mathcal{I}_a \right]$$

$$= \frac{|\mathcal{I}_a| \left(\max_{b \in [n_2]} \|v_b\|^2 \right)}{n_2} \mathrm{Id}$$

$$= \frac{|\mathcal{I}_a| \mu_2^2\left(V\right) r}{n_2^2} \mathrm{Id},$$

where the fourth line is due to the definition of $\mu_2^2\left(V\right)$. This implies that

$$\sigma^2\left(\mathcal{I}_a\right) \leq \frac{|\mathcal{I}_a| \mu_2^2\left(V\right) r}{n_2^2}.$$

20

To find an appropriate $K > 0$ note that almost surely

$$\left\| X_i - \mathbb{E}X_i \right\| = \left\| X_i - \frac{1}{n_1 n_2}\mathrm{Id} \right\|$$

$$\leq \max_{i \in [n_2]}\left\| v_i v_i^* - \frac{1}{n_1 n_2}\mathrm{Id} \right\|$$

$$\leq \frac{1}{n_1 n_2} + \max_{i \in [n_2]}\left\| v_i \right\|^2$$

$$\leq \left(\frac{1}{n_1 r} + 1 \right) \max_{i \in [n_2]}\left\| v_i \right\|^2$$

$$\leq \frac{2r}{n_2}\mu_2^2(V) =: K,$$

where in the fourth line we used that $\max_{i \in [n_2]}\left\| v_i \right\|^2 \geq \frac{r}{n_2} = \frac{1}{n_2}\sum_{i \in [n_2]}\left\| v_i \right\|^2$. Finally to apply Bernstein inequality we need that the X_i's are independent conditionally on $\mathcal{I}_a$, which follows from the fact that the a_i's and b_i's are drawn independently. With these ingredients the matrix Bernstein inequality yields that

$$\mathbb{P}\left(\left\| \sum_{i \in \mathcal{I}_a} X_i - \frac{|\mathcal{I}_a|}{n_2}\mathrm{Id} \right\| \leq t \,\Big|\, \mathcal{I}_a \right) \geq 1 - 2r\exp\left(-c\min\left(\frac{t^2}{\sigma^2(\mathcal{I}_a)}; \frac{t}{K} \right) \right)$$

$$\geq 1 - 2r\exp\left(-c\min\left\{ \frac{n_2^2 t^2}{|\mathcal{I}_a| r \mu_2^2(V)}; \frac{n_2 t}{2r \mu_2^2(V)} \right\} \right).$$

Setting $t = \frac{m}{2n_1 n_2}$ this implies that for fixed $a \in [n_1]$ it holds that

$$\mathbb{P}\left(\left\| \sum_{i \in \mathcal{I}_a} X_i \right\| \leq \frac{m}{2n_1 n_2} + \frac{|\mathcal{I}_a|}{n_2} \,\Big|\, \mathcal{I}_a \right) \geq \mathbb{P}\left(\left\| \sum_{i \in \mathcal{I}_a} X_i - \frac{|\mathcal{I}_a|}{n_2}\mathrm{Id} \right\| \leq \frac{m}{2n_1 n_2} \,\Big|\, \mathcal{I}_a \right)$$

$$\geq 1 - 2r\exp\left(-\frac{cm}{4rn_1\mu_2^2(V)}\min\left\{ \frac{m}{|\mathcal{I}_a| n_1}; 1 \right\} \right). \tag{24}$$

To complete the proof we restrict our attention to $A := \left\{ a \in [n_1] : |\mathcal{I}_a| \leq \frac{4m}{3n_1} \right\}$ as it follows from (23) that

$$\left\{ |\mathcal{I}_a| \leq \frac{4m}{3n_1} \right\} \cap \left\{ \left\| \sum_{i \in \mathcal{I}_a} X_i \right\| \leq \frac{m}{2n_1 n_2} + \frac{|\mathcal{I}_a|}{n_2} \right\} \subset \mathcal{E}_a, \tag{25}$$

and, consequently, for $a \in A$ we obtain that

$$\mathbb{P}\left(\mathcal{E}_a \,\big|\, \mathcal{I}_a \right) \overset{(25),(24),}{\geq} 1 - 2r\exp\left(-\frac{cm}{4rn_1\mu_2^2(V)}\min\left\{ \frac{m}{|\mathcal{I}_a| n_1}; 1 \right\} \right)$$

$$\overset{a \in A}{\geq} 1 - 2r\exp\left(-\frac{3cm}{16rn_1\mu_2^2(V)} \right). \tag{26}$$

As the $\mathcal{E}_a$'s only depend on $\{b_i\}_{i \in \mathcal{I}_a}$ and are hence independent conditionally on $\mathcal{I}_\alpha$, this implies that

$$
\begin{aligned}
\mathbb{P}\left(\bigcap_{a \in A} \mathcal{E}_a^c \,\middle|\, \{\mathcal{I}_a\}_{a=1}^{n_1}\right) &= \prod_{a \in A} \mathbb{P}\left(\mathcal{E}_a^c \,\middle|\, \{\mathcal{I}_a\}_{a=1}^{n_1}\right) \\
&= \prod_{a \in A} \mathbb{P}\left(\mathcal{E}_a^c \,\middle|\, \mathcal{I}_a\right) \\
&\overset{(26)}{\leq} \prod_{a \in A}\left(2r \exp\left(-\frac{3cm}{16rn_1\mu_2^2(V)}\right)\right) \\
&= \left(2r \exp\left(-\frac{3cm}{16rn_1\mu_2^2(V)}\right)\right)^{|A|} \\
&= \left(\exp\left(\log(2r) - \frac{3cm}{16rn_1\mu_2^2(V)}\right)\right)^{|A|} \\
&\leq \left(\exp\left(-\frac{m}{C_1' r n_1 \mu_2^2(V)}\right)\right)^{|A|},
\end{aligned}
$$

where in the last line we have used assumption (21) with C_1 large enough. Furthermore, note that

$$
m = \sum_{a=1}^{n_1} |\mathcal{I}_a| \geq \sum_{i \in [n_1] \setminus A} |\mathcal{I}_a| \geq \frac{4m}{3n_1}\big|\,[n_1]\setminus A\big| = \frac{4m}{3n_1}(n_1 - |A|)
$$

implies that $|A| \geq \frac{n_1}{4}$ almost surely. Hence, it follows that

$$
\mathbb{P}\left(\bigcap_{a \in A} \mathcal{E}_a^c \,\middle|\, \{\mathcal{I}_a\}_{a=1}^{n_1}\right) \leq \exp\left(-\frac{m}{C_2 r \mu_2^2(V)}\right).
$$

This shows that conditional on $\{\mathcal{I}_a\}_{a=1}^{n_1}$ we have that almost surely

$$
\mathbb{P}\left(\bigcup_{a \in [n_1]} \mathcal{E}_a \,\middle|\, \{\mathcal{I}_a\}_{a=1}^{n_1}\right) \geq \mathbb{P}\left(\bigcup_{a \in A} \mathcal{E}_a \,\middle|\, \{\mathcal{I}_a\}_{a=1}^{n_1}\right) \geq 1 - \exp\left(-\frac{m}{C_2 r \mu_2^2(V)}\right).
$$

Taking expectations yields the claim. $\qquad\square$

Now we are prepared to give a proof of Proposition 3.6.

Proof of Proposition 3.6. For the proof we will condition on two events $\mathcal{E}_1$ and $\mathcal{E}_2$, which we will define in the following. The event $\mathcal{E}_1$ is defined by

$$
\mathcal{E}_1 := \left\{\|\mathcal{A}(UV^*)\|^2 \leq 2\|UV^*\|_F^2 = 2r\right\}. \tag{27}
$$

Observe that

$$\|\mathcal{A}(UV^*)\|^2 = \frac{n_1 n_2}{m} \sum_{i=1}^{m} |(UV^*)_{a_i,b_i}|^2 = \sum_{i=1}^{m} X_i,$$

where we have set $X_i := |(UV^*)_{a_i,b_i}|^2$. One has

$$\mathbb{E}X_i = \frac{\|UV^*\|_F^2}{m} = \frac{r}{m},$$

$$\mathbb{E}X_i^2 \leq \frac{r\mu_1^2(X_0)}{m^2}\|UV^*\|_F^2 = \frac{r^2\mu_1^2(X_0)}{m^2}$$

and $X_i \leq \frac{r\mu_1^2(X_0)}{m}$ almost surely, where we used the definition of $\mu_1(X_0)$ and $\|UV^*\|_F^2 = r$. Hence, by the Bernstein inequality (see, e.g., [50, Theorem 2.8.4]) we obtain that

$$\mathbb{P}\left(\left|\sum_{i=1}^{m} X_i - \|UV^*\|_F^2\right| \leq t\right) \leq 2\exp\left(-c\min\left\{\frac{t^2 m^2}{r^2\mu_1^2(X_0)}; \frac{tm}{r\mu_1^2(X_0)}\right\}\right).$$

By setting $t = \|UV^*\|_F^2 = r$ we observe that $\mathcal{E}_1$ holds with probability at least $1 - 2\exp\left(-\frac{cm}{\mu_1^2(X_0)}\right)$.

The second event $\mathcal{E}_2$ is defined by

$$\mathcal{E}_2 := \left\{\exists a \in [n_1] \text{ such that } \|P_{\mathcal{N}_a}V\| \leq \sqrt{\frac{2m}{n_1 n_2}}\right\}.$$

For C_1 in assumption (11) chosen large enough Lemma 4.3 then entails that $\mathbb{P}(\mathcal{E}_2) \geq 1 - \mathcal{O}\left(\exp\left(-\frac{cm}{r\mu_2^2(V)}\right)\right)$. Consequently, we can find $a \in [n_1]$ (depending on the random sampling pattern) such that the condition defining $\mathcal{E}_2$ is satisfied.

Note that in order to prove Proposition 3.6 it is enough to find $Z \in \overline{\mathcal{K}_*(X_0)} \setminus \{0\}$ such that

$$\|\mathcal{A}(Z)\| < 8\sqrt{\frac{m}{rn_1 n_2}}\|Z\|_F, \tag{28}$$

because by definition of the closure and the continuity of $\mathcal{A}$ this implies that there is a matrix $\tilde{Z} \in \mathcal{K}_*(X_0) \setminus \{0\}$ such that

$$\frac{\left\|\mathcal{A}\left(\tilde{Z}\right)\right\|}{\|\tilde{Z}\|_F} \leq 8\sqrt{\frac{m}{rn_1 n_2}},$$

which implies (12) with constant $C_3 = 8$. In the following we will construct such a matrix Z. Let $x \in \mathbb{R}^r$ be a vector such that $\|x\| = 1$. Then for $a \in [n_1]$ as above we define the vector $w_a \in \mathbb{R}^{n_2}$ by

$$w_a := P_{\mathcal{N}_a^\perp}Vx$$

and set

$$W := -\frac{\langle e_a w_a^*, UV^* \rangle_F}{|\langle e_a w_a^*, UV^* \rangle_F|} e_a w_a^* = \frac{\langle e_a w_a^*, UV^* \rangle_F}{|\langle e_a w_a^*, UV^* \rangle_F|} e_a x^* V^* P_{\mathcal{N}_a^\perp}.$$

It follows directly from the definition of $\mathcal{N}_a$ that $\mathcal{A}(W) = 0$. In the following let T be the tangent space of the manifold of rank-r matrices at X_0 as defined in (13). Furthermore, denote by $P_U = UU^*$ the orthogonal projection onto the column space of U and, analogously, by $P_V = VV^*$ the orthogonal projection onto the column space of V. Then we obtain that

$$\begin{aligned}
\|\mathcal{P}_{T^\perp} W\|_* &= \|\mathcal{P}_{T^\perp} (e_i w_a^*)\|_* \\
&= \|P_{U^\perp} e_i w_a^* P_{V^\perp}\|_* \\
&= \|P_{U^\perp} e_a w_a^* P_{V^\perp}\|_F \\
&= \|P_{U^\perp} e_a\| \|P_{V^\perp} w_a\| \\
&\leq \|P_{V^\perp} w_a\|,
\end{aligned}$$

where in the second equality we have used that $\mathcal{P}_{T^\perp} M = P_{U^\perp} M P_{V^\perp}$ for all $M \in \mathbb{R}^{n_1 \times n_2}$ and in the last line we used that $\|P_{U^\perp} e_i\| \leq 1$. Plugging in $w_a = P_{\mathcal{N}_a^\perp} V m$ it follows that

$$\begin{aligned}
\|\mathcal{P}_{T^\perp} W\|_* &\leq \|P_{V^\perp} P_{\mathcal{N}_a^\perp} V x\| \\
&= \|P_{V^\perp} P_{\mathcal{N}_a} V x\|,
\end{aligned}$$

where the last line is due to $P_{V^\perp} V x = 0$. The fact that $\|P_{V^\perp}\| \leq 1$ then yields that

$$\begin{aligned}
\|\mathcal{P}_{T^\perp} W\|_* &\leq \|P_{\mathcal{N}_a} V x\| \\
&\leq \|P_{\mathcal{N}_a} V\| \|x\| \\
&\leq \sqrt{\frac{2m}{n_1 n_2}}
\end{aligned} \tag{29}$$

where for the last line we used that $a \in [n_1]$ was chosen such that the condition in $\mathcal{E}_2$ holds. This shows that W is relatively close to T. Based on W we now aim to find $Z \in \overline{\mathcal{K}_*(X)}$ of the form

$$Z := W - \beta UV^*,$$

where $\beta > 0$ will be chosen in the following such that $Z \in \overline{\mathcal{K}_*(X_0)}$, which by Lemma 4.1 is equivalent to $-\langle UV^*, Z \rangle_F \geq \|\mathcal{P}_{T^\perp} Z\|_*$. First, we note that

$$\|\mathcal{P}_{T^\perp} Z\|_* = \|\mathcal{P}_{T^\perp} W\|_* \leq \sqrt{\frac{2m}{n_1 n_2}} \tag{30}$$

due to $\mathcal{P}_{T^\perp}(UV^*) = 0$ and the inequality chain (29). Furthermore, we have that

$$\begin{aligned}
-\langle UV^*, Z \rangle_F &= r\beta - \langle UV^*, W \rangle_F \\
&= r\beta + \langle UV^*, \frac{\langle e_a w_a^*, UV^* \rangle_F}{|\langle e_a w_a^*, UV^* \rangle_F|} e_a w_a^* \rangle_F \\
&= r\beta + |\langle e_a w_a^*, UV^* \rangle_F| \\
&\geq r\beta.
\end{aligned} \tag{31}$$

24

Hence, setting $\beta = 2\sqrt{\frac{m}{r^2 n_1 n_2}}$ and combining (30) and (31) it follows that $Z \in \overline{\mathcal{K}_*(X_0)}$. This Z also satisfies (28). To see that we observe that

$$
\begin{aligned}
\|Z\|_F &\geq \|e_a w_a^*\|_F - \beta \|UV^*\|_F \\
&= \|w_a\| - \beta\sqrt{r} \\
&= \|P_{\mathcal{N}_a^\perp} Vx\| - \beta\sqrt{r} \\
&= \sqrt{\|Vx\|^2 - \|P_{\mathcal{N}_a} Vx\|^2} - \beta\sqrt{r} \\
&\geq \sqrt{1 - \frac{2m}{n_1 n_2}} - 2\sqrt{\frac{m}{rn_1 n_2}} \\
&> \frac{1}{2},
\end{aligned}
$$

where in the last line we used the assumption that $m \leq \frac{n_1 n_2}{32}$. Furthermore, from $\mathcal{A}(W) = 0$ it follows that

$$
\|\mathcal{A}(Z)\| = \beta\|\mathcal{A}(UV^*)\| \overset{(27)}{\leq} \beta\sqrt{2r} = 2\sqrt{\frac{2m}{rn_1 n_2}} < 8\sqrt{\frac{m}{rn_1 n_2}}\|Z\|_F.
$$

Combining the last two inequality chains implies (28), which completes the proof. $\qquad\square$

4.4. Proof of Theorem 3.1 and Theorem 3.5

As already mentioned Theorem 3.1 can be deduced from Proposition 3.3 and Theorem 3.5 can be deduced from Proposition 3.6. We only show how to prove Theorem 3.1, as the proof of Theorem 3.5 is analogous.

Proof of Theorem 3.1. By Proposition 3.6 and the definition of the minimum conic singular value $\lambda_{\min}(\mathcal{A}, \mathcal{K}_*(h_0 m_0^*))$ with probability at least $1 - \mathcal{O}\left(\exp\left(-\frac{K}{C_2 \mu^2}\right)\right)$ there is a matrix $Z \in \mathcal{K}_*(h_0 m_0^*) \setminus \{0\}$ such that

$$
\|\mathcal{A}(Z)\| \leq C_3 \sqrt{\frac{L}{KN}}\|Z\|_F. \tag{32}
$$

and such that $\tilde{X}_t := h_0 m_0^* + tZ$ obeys $\|\tilde{X}_t\|_* \leq \|h_0 m_0^*\|_*$ for all $0 < t \leq 1$. Next, set $e_t = \frac{t}{2}\mathcal{A}(Z)$. Then for $y_t = \mathcal{A}(h_0 m_0^*) + e_t$ we have that

$$
\|\mathcal{A}\left(\tilde{X}_t\right) - \tilde{y}_t\| = \|\mathcal{A}(tZ) - e\| = \frac{t}{2}\|\mathcal{A}(Z)\|.
$$

Hence, by setting $\tau_0 := \frac{\|\mathcal{A}(Z)\|}{2}$ we observe that $\|\mathcal{A}\left(\hat{X}_t\right) - y_t\| = t\tau_0$. Furthermore, note that

$$
\|\tilde{X}_t - X_0\|_F = \|tZ\|_F \overset{(32)}{\geq} \frac{t}{C_3}\sqrt{\frac{KN}{L}}\|\mathcal{A}(Z)\| = \frac{2t\tau_0}{C_3}\sqrt{\frac{KN}{L}}
$$

Now let $0 < \tau \leq \tau_0$. Then by setting $t = \frac{\tau}{\tau_0}$, $\tilde{X} := \tilde{X}_t$, $y := y_t$, and $e := e_t$ the desired claim follows. $\qquad\square$

5. Stability of blind deconvolution

5.1. Outline of the proof and main ideas

The goal of this section is to prove Theorem 3.7. We first give a proof sketch and present the key ideas. We have seen in Proposition 3.3 that for certain isometries $B \in \mathbb{C}^{L \times K}$ with high probability one has that $\lambda_{\min}\left(\mathcal{A}, \mathcal{K}_*\left(h_0 m_0^*\right)\right) \lesssim \sqrt{\frac{L}{KN}}$. Hence, applying Theorem 2.4 cannot lead to very strong error estimates. However, if we closely inspect the proof of Proposition 3.3 we observe the following. Again, denote by T the tangent space of the manifold of rank-1 matrices at point $h_0 m_0^*$ as defined in (13) and assume that $\|h_0\| = \|m_0\| = 1$. By construction we have that $Z = W - \beta h_0 m_0^*$, where $W = \frac{\langle h_0, e_i \rangle}{\|m_i^\perp\| |\langle h_0, e_i \rangle|} e_i \left(m_i^\perp\right)^*$. This implies that

$$
\begin{aligned}
|\langle Z, h_0 m_0^* \rangle_F| &\leq \left| \langle \frac{\langle h_0, e_i \rangle}{\|m_i^\perp\| |\langle h_0, e_i \rangle|} e_i \left(m_i^\perp\right)^*, h_0 m_0^* \rangle_F \right| + \beta \|h_0 m_0^*\|_F^2 \\
&= |\langle e_i, h_0 \rangle| \|m_i^\perp\| + \beta \|h_0 m_0^*\|_F^2 \\
&\leq \frac{\mu}{\sqrt{L}} + \beta \lesssim \frac{\mu}{\sqrt{L}} + \sqrt{\frac{L}{KN}},
\end{aligned}
$$

where we have used the triangle inequality in the first line and the definition of $m_i^\perp$ in the second line. In the third line we used that $\|m_i^\perp\| \leq 1$, $|\langle e_i, h_0 \rangle| \leq \frac{\mu}{\sqrt{L}}$, and $\beta = \sqrt{\frac{L}{KN}}$. As $\|Z\|_F \gtrsim 1$ this implies that $\frac{\langle Z, h_0 m_0^* \rangle_F}{\|Z\|_F \|h_0 m_0^*\|_F}$ is quite small, meaning that Z and $h_0 m_0^*$ are almost orthogonal to each other. All the descent directions Z with this property, however, have in common that the admissible descent step size

$$
t_0 := \max \left\{ t > 0 : \|h_0 m_0^* + tZ\|_* \leq \|h_0 m_0^*\|_* \right\}
$$

is necessarily very small. Geometrically this corresponds to the fact that the nuclear norm ball is curved near X_0, which is why its near-tangential behaviour only holds locally. This will be made precise in Lemma 5.7 below. For this reason the idea of the proof of Theorem 3.7 is to split the descent cone into two parts. One part will consist of all the matrices aligned with $h_0 m_0^*$. The second part will consist of all remaining matrices, which are almost orthogonal to $h_0 m_0^*$. As mentioned above, these matrices must necessarily be close to T. The first part is captured by the set $\mathcal{E}_{\mu, \delta}$ with

$$
\mathcal{E}_{\mu, \delta} := \bigcup_{h_0 \in \mathcal{H}_\mu, m_0 \in \mathbb{C}^N} \left\{ Z \in \mathcal{K}_*\left(h_0 m_0^*\right) : \delta \leq \frac{-\operatorname{Re}\left(\langle Z, h_0 m_0^* \rangle_F\right)}{\|h_0 m_0^*\|_F} \text{ and } \|Z\|_F = 1 \right\},
$$

where $\delta > 0$. In Section 5.2 we will show that with high probability it holds that

$$
\inf_{Z \in \mathcal{E}_{\mu, \delta}} \|\mathcal{A}(Z)\| \gtrsim \frac{\delta^2}{\log^2(L) \mu^2}. \tag{33}
$$

Hence, if we have for the minimizer $\hat{X}$ of (3) that $\hat{X} - h_0 m_0^*$ is an element of the conic hull of $\mathcal{E}_{\mu, \delta}$ we can proceed similarly as in [9] to obtain near-optimal error bounds. Let us

briefly explain which property of $\mathcal{E}_{\mu,\delta}$ allows us to show (33). We define for any matrix $W \in \mathbb{C}^{K \times N}$ its $\| \cdot \|_{B_1}$-norm by

$$\|W\|_{B_1} := \sum_{\ell=1}^{L} \|W^* b_\ell\|.$$

In other words $\|W\|_{B_1}$ is the ℓ_1-norm of the vector $(\|W^* b_\ell\|)_{\ell=1}^{L}$. We show in Lemma 5.5 below that all $Z \in \mathcal{E}_{\mu,\delta}$ have rather large $\| \cdot \|_{B_1}$-norm, which entails that the mass of the vector $(\|W^* b_\ell\|)_{\ell=1}^{L}$ cannot be concentrated on only very few entries. This in turn will allow us to employ a non-i.i.d. version of Mendelson's smalll bal method [45], allowing us to show a lower bound for (33), see Lemma 5.6 below. To understand the behaviour on the second part recall from Proposition 3.3 that for matrices $Z/\|Z\|_F \in \mathcal{K}_* (h_0 m_0^*) \setminus \mathcal{E}_{\mu,\delta}$ the quantity $\|\mathcal{A}(Z)\|$ may be quite small, so a uniform bound is not feasible. However, as Z is almost orthogonal to $h_0 m_0^*$, also $\|\mathcal{P}_{T^\perp} Z\|_*$ must be rather small because of the characterization of the descent cone, Lemma 4.1. Hence, Z is close to the tangent space and is almost orthogonal to $h_0 m_0^*$. For that reason, whenever $\|h_0 m_0^* + tZ\|_* \leq \|h_0 m_0^*\|_*$ holds, the cylindrical shape of the nuclear norm ball implies that $t > 0$ is small. This fact is captured by Lemma 5.7 below. Theorem 3.3 can then be proven by combining inequality (33) and Lemma 5.7, see Section 5.3.

5.2. An lower bound for the minimum conic singular value

First we recall the notion of Gaussian width (see, e.g., [50]).

Definition 5.1. *For a set $\mathcal{E} \subset \mathbb{C}^{K \times N}$ its Gaussian width is defined by*

$$\omega(\mathcal{E}) := \mathbb{E}\left[\sup_{X \in \mathcal{E}} Re\left(\langle X, G \rangle_F\right)\right],$$

where $G \in \mathbb{C}^{K \times N}$ is a matrix, whose entries are independent and identically distributed random variables with distribution $\mathcal{CN}(0, 1)$.

This definition allows us to state the following lemma, which is important for our analysis of the conic singular value. It relies on a uniform lower bound on the number of measurements whose magnitude is larger than a certain constant and is a variant of Theorem 2.1 in [45].

Lemma 5.2. *Let $\mathcal{E} \subset \mathbb{C}^{K \times N}$ be a symmetric set, i.e., $\mathcal{E} = -\mathcal{E}$. For all $\xi > 0$ and $t > 0$ it holds with probability at least $1 - \exp\left(-2t^2\right)$ that*

$$\inf_{X \in \mathcal{E}} \left| \{\ell \in [L]: \ |\langle b_\ell c_\ell^*, X \rangle_F| \geq \xi\} \right| \geq \inf_{X \in \mathcal{E}} \left(\sum_{\ell=1}^{L} \mathbb{P}\left(|\langle b_\ell c_\ell^*, X \rangle_F| \geq 2\xi\right) \right) - \frac{4\omega(\mathcal{E})}{\xi} - t\sqrt{L}.$$

$$(34)$$

Here $\varepsilon_1, \ldots, \varepsilon_L$ are independent Rademacher variables, i.e., random variables which take the two values ± 1 each with probability $\frac{1}{2}$.

The proof of the Lemma 5.2 is based on a variant of Mendelson's small-ball method and proceeds in analogy to [45]. We have deferred a detailed proof to Appendix A. In order to apply Lemma 5.2 we need to estimate the first term of the right-hand side of Lemma 5.2. Such an estimate can be derived using the Paley-Zygmund inequality as in [45]. For the sake of completeness we have included a proof in Appendix B.

Lemma 5.3. *Let $X \in \mathbb{C}^{K \times N}$ be an arbitrary matrix. Then for all $\xi > 0$*

$$\sum_{\ell=1}^{L} \mathbb{P}\left(|\langle b_\ell c_\ell^*, X \rangle_F| \geq 2\xi\right) \geq \frac{9}{32} \big| \{\ell \in [L] : \ \|X^* b_\ell\| \geq 4\xi\} \big|.$$

In order to use Lemma 5.3 we need a lower bound for $|\{\ell \in [L] : \ \|X^* b_\ell\| \geq \xi\}|$. This will be achieved by the next lemma. For the statement of this lemma and its proof we will need to introduce the following notion. We define for any matrix $W \in \mathbb{C}^{K \times N}$ its $\|\cdot\|_{B_{1,w}}$-quasinorm by

$$\|W\|_{B_{1,w}} := \sup_{\xi \geq 0} \xi \big| \{\ell \in [L] : \|W^* b_\ell\| \geq \xi\} \big|.$$

That is, $\|W\|_{B_{1,w}}$ is the weak ℓ_1-norm of the vector $(\|W^* b_\ell\|)_{\ell=1}^{L}$.[3] A direct consequence of this interpretation is the inequality (see, e.g., [53, Proposition 2.10 and Exercise 2.4])

$$\|W\|_{B_{1,w}} \leq \|W\|_{B_1} \leq \log(eL) \|W\|_{B_{1,w}}. \tag{35}$$

Lemma 5.4. *Let $Z \in \mathbb{C}^{K \times N}$ such that $\|Z\|_F = 1$. Then it holds that*

$$\left| \left\{ \ell \in [L] : \ \|Z^* b_\ell\| \geq \frac{\|Z\|_{B_1}}{L \log(eL)} \right\} \right| \geq \frac{\|Z\|_{B_1}^2}{\log^2(eL)}.$$

Proof of Lemma 5.4. Choose ξ^* such that

$$\|Z\|_{B_{1,w}} = \xi^* |\{\ell \in [L] : \ \|Z^* b_\ell\| \geq \xi^*\}|. \tag{36}$$

As $|\{\ell \in [L] : \ \|Z^* b_\ell\| \geq \xi^*\}| \leq L$ it follows that

$$\xi^* \geq \frac{\|Z\|_{B_{1,w}}}{L} \geq \frac{\|Z\|_{B_1}}{L \log(eL)}, \tag{37}$$

where we also used inequality (35). We observe that

$$
\begin{aligned}
1 = \|Z\|_F^2 \\
= \sum_{\ell=1}^{L} \|Z^* b_\ell\|^2 \\
\geq \xi^{*2} |\{\ell \in [L] : \ \|Z^* b_\ell\| \geq \xi^*\}| \\
\overset{(36)}{=} \frac{\|Z\|_{B_{1,w}}^2}{|\{\ell \in [L] : \ \|Z^* b_\ell\| \geq \xi^*\}|} \\
\overset{(35)}{\geq} \frac{\|Z\|_{B_1}^2}{\log^2(eL) |\{\ell \in [L] : \ \|Z^* b_\ell\| \geq \xi^*\}|},
\end{aligned}
$$

---

[3]For a more detailed discussion of the weak ℓ_1-norm see, e.g., [52].

where for the second equality we used the identity $\sum_{\ell=1}^{L} b_\ell b_\ell^* = \text{Id}$. Using inequality (37) and rearranging terms it follows that

$$\left| \left\{ \ell \in [L] : \ \|Z^* b_\ell\| \geq \frac{\|Z\|_{B_1}}{L \log(eL)} \right\} \right| \geq \left| \left\{ \ell \in [L] : \ \|Z^* b_\ell\| \geq \xi^* \right\} \right| \geq \frac{\|Z\|_{B_1}^2}{\log^2(eL)},$$

which completes the proof.

$\square$

The next lemma gives a bound on $\inf_{Z \in \mathcal{E}_{\mu,\delta}} \|Z\|_{B_1}$.

Lemma 5.5. *It holds that*

$$\inf_{Z \in \mathcal{E}_{\mu,\delta}} \|Z\|_{B_1} \geq \frac{\delta \sqrt{L}}{\mu}.$$

Proof. Let $Z \in \mathcal{E}_{\mu,\delta}$ be arbitrary. By definition of $\mathcal{E}_{\mu,\delta}$ there is $h_0 \in \mathcal{H}_\mu$ and $m_0 \in \mathbb{C}^N$ such that $Z \in \mathcal{K}_* (h_0 m_0^*)$ and such that the inequality

$$\delta \leq \frac{-\text{Re}\left(\langle Z, h_0 m_0^* \rangle_F\right)}{\|h_0 m_0^*\|_F}$$

holds. It follows that

$$\begin{aligned}
\delta &\leq \frac{-\text{Re}\left(\langle Z m_0, h_0 \rangle\right)}{\|h_0 m_0^*\|_F} \\
&= \frac{-\sum_{\ell=1}^{L} \text{Re}\left(\langle Z m_0, b_\ell \rangle \langle b_\ell, h_0 \rangle\right)}{\|h_0 m_0^*\|_F} \\
&\leq \frac{\left(\max_{\ell \in [L]} |\langle h_0, b_\ell \rangle|\right) \sum_{\ell=1}^{L} |\langle Z m_0, b_\ell \rangle|}{\|h_0 m_0^*\|_F},
\end{aligned}$$

where for the second equality we have used that $\sum_{\ell=1}^{L} b_\ell b_\ell^* = \text{Id}$. Note that for all $\ell \in [L]$ it holds that

$$\frac{|\langle Z m_0, b_\ell \rangle|}{\|m_0\|} \leq \|Z^* b_\ell\|.$$

Hence, by the previous inequality chain it follows that

$$\delta \leq \frac{\max_{\ell \in [L]} |\langle h_0, b_\ell \rangle|}{\|h_0\|} \sum_{\ell=1}^{L} \|Z^* b_\ell\| \leq \frac{\mu}{\sqrt{L}} \|Z\|_{B_1},$$

where in the last inequality we used the definition of μ and $\|Z\|_{B_1}$. Rearranging terms and taking the infimum over all $Z \in \mathcal{E}_{\mu,\delta}$ yields the desired inequality. $\square$

Having gathered all the necessary ingredients we can state and prove the main lemma of this section.

Lemma 5.6. *Let $\delta > 0$. Assume that*

$$L \geq C_1 \left(\frac{\mu}{\delta} \right)^6 (K + N) \, \log^6 (eL), \tag{38}$$

Then with probability at least $1 - \exp \left(-\frac{L\delta^4}{C_2 \log^4 (eL) \mu^4} \right)$ it holds that

$$\inf_{Z \in \mathcal{E}_{\mu,\delta}} \| \mathcal{A}(Z) \| \gtrsim \frac{\delta^2}{\log^2 (L) \, \mu^2}. \tag{39}$$

C_1 and C_2 are absolute constants.

Proof. Our goal is to apply Lemma 5.2. In order to apply it we first derive a lower bound for the first term on the right-hand side of inequality (34). For that recall that by Lemma 5.5 it holds that

$$\inf_{Z \in \mathcal{E}_{\mu,\delta}} \| Z \|_{B_1} \geq \frac{\delta \sqrt{L}}{\mu}. \tag{40}$$

Thus, for any $Z \in \mathcal{E}_{\mu,\delta}$ we obtain that

$$
\left| \left\{ \ell \in [L] : \ \| Z^* b_\ell \| \geq \frac{\delta}{\mu \sqrt{L} \log (eL)} \right\} \right| \geq \left| \left\{ \ell \in [L] : \ \| Z^* b_\ell \| \geq \frac{\| Z \|_{B_1}}{L \log (eL)} \right\} \right|
$$
$$
\geq \frac{\| Z \|_{B_1}^2}{\log^2 (eL)}
$$
$$
\geq \frac{\delta^2 L}{\mu^2 \log^2 (eL)},
$$

where the first inequality follows from (40), the second one is due to Lemma 5.4, and the third one follows again from (40). Hence, by Lemma 5.3 applied with $\xi = \frac{\delta}{4 \sqrt{L} \ln(eL) \mu}$ we finally obtain that

$$
\begin{aligned}
\inf_{Z \in \mathcal{E}_{\mu,\delta}} &\left(\sum_{\ell=1}^{L} \mathbb{P} \left(|\langle b_\ell c_\ell^*, Z \rangle_F| \geq \frac{\delta}{2 \sqrt{L} \ln (eL) \, \mu} \right) \right) \\
&\geq \frac{9}{32} \inf_{Z \in \mathcal{E}_{\mu,\delta}} \left| \left\{ \ell \in [L] : \ \| Z^* b_\ell \| \geq \frac{\delta}{\sqrt{L} \ln (eL) \, \mu} \right\} \right| \\
&\geq \frac{9 \delta^2 L}{32 \mu^2 \log^2 (eL)}.
\end{aligned} \tag{41}
$$

Next, we need an upper bound for the Gaussian width. For that, we first observe that

$$\mathcal{E}_{\mu,\delta} \subset \left(\bigcup_{h_0 \in \mathbb{C}^K, m_0 \in \mathbb{C}^N} \mathcal{K}_* (h_0 m_0^*) \right) \cap \left\{ Z \in \mathbb{C}^{K \times N} : \ \| Z \|_F = 1 \right\} =: \mathcal{E}.$$

The Gaussian width of $\mathcal{E}$ has been bounded in [44, Lemma 4.1], combined with the monotonicity of the Gaussian width their results yields that

$$\omega (\mathcal{E}_{\mu,\delta}) \leq \omega (\mathcal{E}) \leq 2 \sqrt{(K + N)}. \tag{42}$$

Thus for $\xi = \frac{\delta}{4\sqrt{L}\log(eL)\mu}$ we obtain from Lemma 5.2 together with (41), (42) that with probability at least $1 - \exp\left(-2t^2\right)$ it holds that

$$\inf_{Z \in \mathcal{E}_{\mu,\delta}} \left| \{\ell \in [L] : \ |\langle b_\ell c_\ell^*, Z\rangle_F| \geq \xi\} \right| \geq \frac{9L\delta^2}{32\log^2(eL)\mu^2} - \frac{2\log(eL)\mu\sqrt{L(K+N)}}{\delta} - t\sqrt{L}$$

$$\geq \frac{9L\delta^2}{64\log^2(eL)\mu^2} - t\sqrt{L},$$

where the second inequality follows from assumption (38), if the constant $C_1 > 0$ is chosen large enough. Consequently, setting $t = \frac{9\delta^2\sqrt{L}}{128\log^2(eL)\mu^2}$ and recalling that $(\mathcal{A}(Z))(\ell) = \langle b_\ell c_\ell^*, Z\rangle_F$ we have that with probability at least $1 - \exp\left(-\frac{L\delta^4}{C_2\log^4(eL)\mu^4}\right)$ with C_2 chosen appropriately

$$\inf_{Z \in \mathcal{E}_{\mu,\delta}} \left| \left\{\ell \in [L] : \ |\mathcal{A}(Z)(\ell)| \geq \frac{\delta}{4\sqrt{L}\log(eL)\mu}\right\} \right| \geq \frac{9L\delta^2}{128\log^2(L)\mu^2}.$$

Summing up we obtain that with probability at least $1 - \exp\left(-\frac{L\delta^4}{C_2\log^4(eL)\mu^4}\right)$

$$\inf_{Z \in \mathcal{E}_{\mu,\delta}} \|\mathcal{A}(Z)\| \geq \inf_{Z \in \mathcal{E}_{\mu,\delta}} \frac{\delta}{4\sqrt{L}\log(eL)\mu} \sqrt{\left|\left\{\ell \in [L] : \ |\mathcal{A}(Z)(\ell)| \geq \frac{\delta}{4\sqrt{L}\log(eL)\mu}\right\}\right|}$$

$$\gtrsim \frac{\delta^2}{\log^2(L)\mu^2}.$$

This shows the claim. $\qquad\square$

5.3. Proof of Theorem 3.7

As already mentioned in Section 4, in order to control all matrices $Z \in \mathcal{K}_*(h_0 m_0^*)$, which are almost orthogonal to $h_0 m_0^*$, we need the following key lemma.

Lemma 5.7. *Let $h_0 m_0^* \in \mathbb{C}^{n_1 \times n_2}$ be a rank-1 matrix. Assume that $Z \in \mathcal{K}_*(h_0 m_0^*) \setminus \{0\}$. Then, whenever $\|h_0 m_0^* + Z\|_* \leq \|h_0 m_0^*\|_*$, it holds that*

$$\|Z\|_F \leq -2Re\left(\langle h_0 m_0^*, \frac{1}{\|Z\|_F}Z\rangle_F\right).$$

Proof. We observe that

$$\|h_0 m_0^*\|_F^2 = \|h_0 m_0^*\|_*^2$$
$$\geq \|h_0 m_0^* + Z\|_*^2$$
$$\geq \|h_0 m_0^* + Z\|_F^2$$
$$= \|h_0 m_0^*\|_F^2 + \|Z\|_F^2 + 2\text{Re}\left(\langle h_0 m_0^*, Z\rangle_F\right).$$

Rearranging terms yields the result. $\qquad\square$

Now we have gathered all tools which are needed to prove Theorem 3.7.

Proof of Theorem 3.7. Having introduced all necessary tools in the last two sections we can now give a proof of Theorem 3.7. We set $\delta := (\log eL)^{2/3} \mu^{2/3} \alpha^{1/3}$. Throughout the proof we will assume that inequality (39) holds, which by Lemma 5.6 holds with probability at least

$$1 - \exp\left(-\frac{L\delta^4}{C_2 \log^4 (eL) \mu^4}\right) = 1 - \exp\left(-\frac{L\alpha^{4/3}}{C_2 \log^{4/3} (eL) \mu^{4/3}}\right).$$

Let $h_0 \in \mathcal{H}_\mu$ and $m_0 \in \mathbb{C}^N$. Furthermore, let $\hat{X}$ be a minimizer of (3) and set $Z := \hat{X} - h_0 m_0^*$. Note that from the minimality of $\hat{X}$ it follows that $\|\hat{X}\|_* \leq \|h_0 m_0^*\|_*$. This implies that $Z \in \mathcal{K}_* (h_0 m_0^*)$. To prove the lemma it remains to derive an appropriate upper bound on $\|Z\|_F$. For that we will distinguish two cases, namely $\frac{Z}{\|Z\|_F} \in \mathcal{E}_{\mu,\delta}$ and $\frac{Z}{\|Z\|_F} \notin \mathcal{E}_{\mu,\delta}$. If $\frac{Z}{\|Z\|_F} \in \mathcal{E}_{\mu,\delta}$, it follows from inequality (39) that

$$
\begin{aligned}
\|Z\|_F &\lesssim \frac{\log^2 (L) \mu^2}{\delta^2} \left\|\mathcal{A}(Z)\right\| \\
&\leq \frac{\log^2 (L) \mu^2}{\delta^2} \left(\|\mathcal{A}(\hat{X}) - y\| + \|e\|\right) \\
&\leq 2\frac{\log^2 (L) \mu^2}{\delta^2} \tau \\
&= 2\frac{\log^{2/3} (L) \mu^{2/3}}{\alpha^{2/3}} \tau,
\end{aligned}
\tag{43}
$$

where in the second inequality we used the triangle inequality as well as $Z = \hat{X} - h_0 m_0^*$ and $y = \mathcal{A}(h_0 m_0^*) + e$. In the third inequality we used that $\hat{X}$ is feasible and $\|e\| \leq \tau$. If $\frac{Z}{\|Z\|_F} \notin \mathcal{E}_{\mu,\delta}$, it follows directly from the definition of $\mathcal{E}_{\mu,\delta}$ that $-\mathrm{Re}\left(\langle \frac{h_0 m_0^*}{\|h_0 m_0^*\|_F}, \frac{Z}{\|Z\|_F}\rangle_F\right) < \delta$. By Lemma 5.7 we obtain that

$$
\begin{aligned}
\|Z\|_F &\leq -2\mathrm{Re}\left(\langle h_0 m_0^*, \frac{1}{\|Z\|_F} Z\rangle_F\right) \\
&< 2\delta \|h_0 m_0^*\|_F \\
&< 2 (\log L)^{2/3} \mu^{2/3} \alpha^{1/3} \|h_0 m_0^*\|_F.
\end{aligned}
\tag{44}
$$

Combining the estimates (43) and (44) we obtain that

$$\|\hat{X} - h_0 m_0^*\|_F = \|Z\|_F \lesssim \frac{\mu^{2/3} \log^{2/3} L}{\alpha^{2/3}} \max\{\tau; \alpha\|h_0 m_0^*\|_F\}.$$

which completes the proof. $\qquad\square$

6. Outlook

In this paper we have analyzed two important cases of structured low-rank matrix recovery problems, blind deconvolution and matrix completion, through an inspection of the descent cone of the nuclear norm and its interaction with the measurement operator $\mathcal{A}$. We have shown that the conic singular value is typically quite small and, consequently, previous analysis approaches cannot give strong recovery guarantees. For the example of blind deconvolution we have presented a new approach based on a refined analysis of the descent cone, showing that the nuclear norm minimization approach is stable against adversarial noise in certain important parameter regimes and allows for uniform recovery guarantees in the presence of noise. In our opinion our results give rise to a number of interesting follow-up questions.

- **Stability for small noise-levels:** Until now, our stability result only covers the situation that the noise level τ is of constant order (up to logarithmic factors). For small τ, Theorem 3.1, respectively Theorem 3.5, put some barriers on what performance can be expected. Nevertheless, it will be interesting to examine the transitional case, that τ is rather small, even further. For example, while the bad conditioning for small noise levels has been established, it remains open whether one can construct a noise vector e such that the true minimizer behaves like the alternative (but non-minimal) solutions constructed in Theorems 3.1 and 3.5. Also the transitional behavior of the minimum conic singular values between very small noise levels (where we established bad conditioning) and larger noise levels (where at least for randomized blind deconvolution, we proved stability) will be an interesting question to study.

- **Extension to the rank r case:** Understanding nuclear norm recovery for matrix completion under adversarial noise remains an important open problem in the field. While our result established that recovery guarantees for arbitrary noise levels are not feasible, our considerations for the rank one scenario give hope that for sufficiently large noise levels, near optimal guarantees are within reach also for matrices of arbitrary rank.

  Similarly, a natural generalization of blind deconvolution is the problem of blind demixing [12, 20], where one observes a noisy superposition of several convolutions, that is, $y = \sum_{i=1}^{r} w_i * x_i + e$. The corresponding low-rank matrix formulation can be interpreted as a rank r version of the randomized blind deconvolution problem.

  We expect that a rank r version of Theorem 3.7 will apply to both these scenarios, which is why we consider this a very promising direction for future research.

- **Extension to other low-rank matrix recovery models**: Various other low-rank matrix models also involve incoherence in some way, for example, robust PCA ([5]) and spectral compressed sensing via matrix completion [54]. Also for these problems, recovery results are typically proven via the Golfing Scheme and lead to a seemingly suboptimal noise bound (see, e.g., [55, Section VI]). Can these problems be analyzed with the methods developed in this paper?

Moreover, [56] provided an incoherence based analysis of the phase retrieval problem under random Bernoulli measurements. It will be interesting to analyze this setup with similar methods as in this manuscript.

Acknowledgements

References

[1] Felix Krahmer and Dominik Stöger. Blind deconvolution: Convex geometry and noise robustness. In *52nd Annual Asilomar Conference on Signals, Systems, and Computers*, 2018.

[2] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717, 2009.

[3] Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Comm. Pure Appl. Math.*, 66(8):1241–1274, 2013.

[4] Ali Ahmed, Benjamin Recht, and Justin Romberg. Blind deconvolution using convex programming. *IEEE Trans. Inform. Theory*, 60(3):1711–1732, 2014.

[5] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[6] Michael B. McCoy and Joel A. Tropp. Sharp recovery bounds for convex demixing, with applications. *Found. Comput. Math.*, 14(3):503–567, 2014.

[7] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.

[8] Yi-Kai Liu. Universal low-rank matrix recovery from pauli measurements. In *Advances in Neural Information Processing Systems*, pages 1638–1646, 2011.

[9] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[10] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15):150401, 2010.

[11] David Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory*, 57(3):1548–1566, 2011.

[12] Shuyang Ling and Thomas Strohmer. Blind deconvolution meets blind demixing: Algorithms and performance bounds. *IEEE Trans. Inform. Theory*, 63(7):4497–4520, 2017.

[13] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Compressed sensing and best k-term approximation. *J. Am. Math. Soc.*, 22(1):211–231, 2009.

[14] Massimo Fornasier, Holger Rauhut, and Rachel Ward. Low-rank matrix recovery via iteratively reweighted least squares minimization. *SIAM J. Optim.*, 21(4):1614–1640, 2011.

[15] Christian Kümmerle and Juliane Sigl. Harmonic mean iteratively reweighted least squares for low-rank matrix recovery. *J. Mach. Learn. Res.*, 19:49, 2018.

[16] Richard Kueng, Huangjun Zhu, and David Gross. Low rank matrix recovery from clifford orbits. *arXiv preprint arXiv:1610.08070*, 2016.

[17] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Appl. Comput. Harmon. Anal.*, 42(1):88–116, 2017.

[18] Yaniv Plan and Roman Vershynin. The generalized lasso with non-linear observations. *IEEE Trans. Inf. Theory*, 62(3):1528–1537, 2016.

[19] Joel A. Tropp. Convex recovery of a structured signal from independent random linear measurements. In *Sampling theory, a renaissance. Compressive sensing and other developments*, pages 67–101. Cham: Birkhäuser/Springer, 2015.

[20] Peter Jung, Felix Krahmer, and Dominik Stöger. Blind demixing and deconvolution at near-optimal rate. *IEEE Trans. Inform. Theory*, 64(2):704–727, 2018.

[21] Michael Kech and Felix Krahmer. Optimal injectivity conditions for bilinear inverse problems with applications to identifiability of deconvolution problems. *SIAM J. Appl. Algebra Geom.*, 1(1):20–37, 2017.

[22] Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *Appl. Comput. Harmon. Anal.*, 2018.

[23] Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. *arXiv preprint arXiv:1711.10467*, 2017.

[24] Wen Huang and Paul Hand. Blind deconvolution by a steepest descent algorithm on a quotient manifold. *SIAM J. Imaging Sci.*, 11(4):2757–2785, 2018.

[25] Jasson D. M. Rennie and Nathan Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 713–719, New York, NY, USA, 2005. ACM.

[26] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. Convex multi-task feature learning. *Machine Learning*, 73(3):243–272, 2008.

[27] Emmanuel J Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2010.

[28] Benjamin Recht. A simpler approach to matrix completion. *J. Mach. Learn. Res.*, 12(Dec):3413–3430, 2011.

[29] Emmanuel J Candès and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

[30] Vladimir Koltchinskii, Karim Lounici, Alexandre B Tsybakov, et al. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.*, 39(5):2302–2329, 2011.

[31] Sahand Negahban and Martin J Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.*, 13(May):1665–1697, 2012.

[32] Olga Klopp. Noisy low-rank matrix completion with general sampling distribution. *Bernoulli*, 20(1):282–303, 2014.

[33] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Trans. Inf. Theory*, 56(6):2980–2998, 2010.

[34] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 665–674, New York, NY, USA, 2013. ACM.

[35] Moritz Hardt. Understanding alternating minimization for matrix completion. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 651–660. IEEE, 2014.

[36] Ruoyu Sun and Zhi-Quan Luo. Guaranteed matrix completion via non-convex factorization. *IEEE Trans. Inform. Theory*, 62(11):6535–6579, 2016.

[37] Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems*, pages 2973–2981, 2016.

[38] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from noisy entries. *J. Mach. Learn. Res.*, 11:2057–2078, 2010.

[39] Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *arXiv preprint arXiv:1902.07698*, 2019.

[40] Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Commun. Pure Appl. Math.*, 61(8):1025–1045, 2008.

[41] Dennis Amelunxen, Martin Lotz, Michael B. McCoy, and Joel A. Tropp. Living on the edge: phase transitions in convex programs with random data. *Inf. Inference*, 3(3):224–294, 2014.

[42] Guillaume Lecué and Shahar Mendelson. Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc. (JEMS)*, 19(3):881–904, 2017.

[43] Sjoerd Dirksen, Guillaume Lecué, and Holger Rauhut. On the gap between restricted isometry properties and sparse recovery conditions. *IEEE Trans. Inform. Theory*, 64(8):5478–5487, 2018.

[44] Maryia Kabanava, Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Stable low-rank matrix recovery via null space properties. *Inf. Inference*, 5(4):405–441, 2016.

[45] Vladimir Koltchinskii and Shahar Mendelson. Bounding the smallest singular value of a random matrix without concentration. *Int. Math. Res. Not. IMRN*, 2015(23):12991–13008, 2015.

[46] Shahar Mendelson. Learning without concentration. In *Conference on Learning Theory*, pages 25–39, 2014.

[47] G.A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra Appl.*, 170:33–45, 1992.

[48] R.Tyrrell Rockafellar. *Convex analysis.* 1970.

[49] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge University Press, 2004.

[50] Roman Vershynin. High-dimensional probability.

[51] Joel A Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[52] Loukas Grafakos. *Classical Fourier analysis. 2nd ed.* New York, NY: Springer, 2nd ed. edition, 2008.

[53] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Birkhäuser Basel, 2013.

[54] Yuxin Chen and Yuejie Chi. Robust spectral compressed sensing via structured matrix completion. *IEEE Trans. Inf. Theory*, 60(10):6576–6601, 2014.

[55] Zihan Zhou, Xiaodong Li, John Wright, Emmanuel J Candès, and Yi Ma. Stable principal component pursuit. In *2010 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 1518–1522. IEEE, 2010.

[56] Felix Krahmer and Yi-Kai Liu. Phase retrieval without small-ball probability assumptions. *IEEE Trans. Inform. Theory*, 64(1):485–500, 2018.

[57] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[58] Aad W Van Der Vaart and Jon A Wellner. Weak convergence. In *Weak Convergence and Empirical Processes*, pages 16–28. Springer, 1996.

[59] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

[60] Víctor de la Peña and Evarist Giné. *Decoupling. From dependence to independence. Randomly stopped processes, U-statistics and processes, martingales and beyond.* New York, NY: Springer, 1998.

A. Proof of Lemma 5.2

The proof of Lemma 5.2 will rely on the following two lemmas. The first lemma is a version of Mendelson's small-ball method for non-i.i.d. measurements. In order to state it let $X_1, \ldots, X_L$ be independent, matrix-valued random variables defined on a probability space (Ω, μ). For every measurable, real-valued function f and for every $\xi > 0$ we define the quantity

$$Q_\xi (f) = \sum_{\ell=1}^{L} \mathbb{P} (f (X_\ell) \geq \xi).$$

Lemma A.1. *Let $X_1, \ldots, X_L \in \mathbb{C}^{K \times N}$ be independent random variables and $\mathcal{F}$ be a set of real-valued functions, which are measurable with respect to (Ω, μ). Let $t > 0$ and $\xi > 0$. Then with probability at least $1 - \exp\left(-2t^2\right)$ it holds that*

$$\inf_{f \in \mathcal{F}} \left| \{\ell \in [L] : f (X_\ell) \geq \xi\} \right| \geq \inf_{f \in \mathcal{F}} Q_{2\xi} (f) - \frac{2}{\xi} \mathbb{E} \left[\sup_{f \in \mathcal{F}} \sum_{\ell=1}^{L} \varepsilon_\ell f (X_\ell) \right] - t\sqrt{L},$$

where $\varepsilon_1, \ldots, \varepsilon_L$ are independent Rademacher variables, i.e., random variables which take the two values ± 1 each with probability $\frac{1}{2}$.

The proof of Lemma A.1 is exactly analogous as the proof of the original small-ball method [45], see Section A.1. The second auxiliary lemma, proved in Section A.2, relates the quantity $\mathbb{E}\left[\sup_{f\in\mathcal{F}}\sum_{\ell=1}^{L}\varepsilon_{\ell}f\left(X_{\ell}\right)\right]$ in the blind deconvolution framework to the Gaussian width (cf. Definition 5.1).

Lemma A.2. *Let $\mathcal{E}\subset\mathbb{C}^{K\times N}$. Then it holds that*

$$\mathbb{E}\left[\sup_{X\in\mathcal{E}}Re\left(\sum_{\ell=1}^{L}b_{\ell}^{*}Xc_{\ell}\right)\right]=\omega\left(\mathcal{E}\right).$$

With these lemmas we can now prove Lemma 5.2.

Proof of Lemma 5.2. Set $X_{\ell}:=b_{\ell}c_{\ell}^{*}$ for all $\ell\in[L]$ and define

$$\mathcal{F}:=\{|\langle M,\cdot\rangle_{F}|:M\in\mathcal{E}\}.$$

Then, by a direct application of Lemma A.1 we obtain that with probability at least $1-\exp\left(-2t^{2}\right)$ it holds that

$$
\inf_{M\in\mathcal{E}}\left|\left\{\ell\in[L]:\ |\langle M,b_{\ell}c_{\ell}^{*}\rangle_{F}|\geq\xi\right\}\right|
$$
$$
\geq\inf_{M\in\mathcal{E}}\left(\sum_{\ell=1}^{L}\mathbb{P}\left(|\langle b_{\ell}c_{\ell}^{*},M\rangle_{F}|\geq2\xi\right)\right)-\frac{2}{\xi}\mathbb{E}\left[\sup_{M\in\mathcal{E}}\sum_{\ell=1}^{L}\varepsilon_{\ell}|\langle b_{\ell}c_{\ell}^{*},M\rangle_{F}|\right]-t\sqrt{L}. \tag{45}
$$

To bound the second summand, we observe that

$$
\mathbb{E}\left[\sup_{M\in\mathcal{E}}\sum_{\ell=1}^{L}\varepsilon_{\ell}|\langle b_{\ell}c_{\ell}^{*},M\rangle_{F}|\right]
$$
$$
\leq\mathbb{E}\left[\sup_{M\in\mathcal{E}}\sum_{\ell=1}^{L}\varepsilon_{\ell}|\text{Re}\left(\langle b_{\ell}c_{\ell}^{*},M\rangle_{F}\right)|\right]+\mathbb{E}\left[\sup_{M\in\mathcal{E}}\sum_{\ell=1}^{L}\varepsilon_{\ell}|\text{Im}\left(\langle b_{\ell}c_{\ell}^{*},M\rangle_{F}\right)|\right]
$$
$$
=2\,\mathbb{E}\left[\sup_{M\in\mathcal{E}}\sum_{\ell=1}^{L}\varepsilon_{\ell}|\text{Re}\left(\langle b_{\ell}c_{\ell}^{*},M\rangle_{F}\right)|\right] \tag{46}
$$
$$
=2\,\mathbb{E}\left[\sup_{M\in\mathcal{E}}\sum_{\ell=1}^{L}\varepsilon_{\ell}\text{Re}\left(\langle b_{\ell}c_{\ell}^{*},M\rangle_{F}\right)\right]
$$
$$
=2\omega\left(\mathcal{E}\right)
$$

where in the third line we used that $\text{Re}\left(\langle b_{\ell}c_{\ell}^{*},X\rangle_{F}\right)$ and $\text{Im}\left(\langle b_{\ell}c_{\ell}^{*},X\rangle_{F}\right)$ have the same distribution. The fourth line follows from the symmetry of the set $\mathcal{E}$ and the last line is due to Lemma A.2. Combining (45) and (46) finishes the proof. $\square$

A.1. Proof of Lemma A.1

We directly trace the steps of the proof of Theorem 1.5 in [45]. In the following $\mathbb{1}_A$ denotes the indicator function, which takes the value 1, if the event A occurs and the value 0 otherwise. Note that

$$\xi \left| \{\ell \in [L]: \ f(X_\ell) \geq \xi\} \right| = \xi \sum_{\ell=1}^{L} \mathbb{1}_{\{f(X_\ell) \geq \xi\}}.$$

Taking the infimum we observe that by the definition of $Q_{2\xi}$

$$
\begin{aligned}
&\xi \inf_{f \in \mathcal{F}} \left| \{\ell \in [L]: \ f(X_\ell) \geq \xi\} \right| \\
&\geq \xi \inf_{f \in \mathcal{F}} Q_{2\xi}(f) - \xi \sup_{f \in \mathcal{F}} \sum_{\ell=1}^{L} \left(\mathbb{P}(f(X_\ell) \geq 2\xi) - \mathbb{1}_{\{f(X_\ell) \geq \xi\}} \right).
\end{aligned}
\tag{47}
$$

The bounded difference inequality (see, for example, [57]) implies that with probability at least $1 - \exp\left(-2t^2\right)$ it holds that

$$
\begin{aligned}
&\sup_{f \in \mathcal{F}} \sum_{\ell=1}^{L} \left(\mathbb{P}(f(X_\ell) \geq 2\xi) - \mathbb{1}_{\{f(X_\ell) \geq \xi\}} \right) \\
&\leq \mathbb{E}\left[\sup_{f \in \mathcal{F}} \sum_{\ell=1}^{L} \left(\mathbb{P}(f(X_\ell) \geq 2\xi) - \mathbb{1}_{\{f(X_\ell) \geq \xi\}} \right) \right] + t\sqrt{L} \\
&= \mathbb{E}\left[\sup_{f \in \mathcal{F}} \sum_{\ell=1}^{L} \left(\mathbb{E}\left[\mathbb{1}_{\{f(X_\ell) \geq 2\xi\}} \right] - \mathbb{1}_{\{f(X_\ell) \geq \xi\}} \right) \right] + t\sqrt{L}
\end{aligned}
\tag{48}
$$

To deal with the expectation we will use the function $\Psi_\xi : [0, +\infty) \longrightarrow \mathbb{R}$ defined by

$$
\Psi_\xi(u) = \begin{cases} 0 & 0 \leq u \leq \xi \\ \frac{1}{\xi}(u - \xi) & \xi \leq u \leq 2\xi \ . \\ 1 & u \geq 2\xi \end{cases}
$$

We observe that Ψ_ξ is Lipschitz continuous with Lipschitz constant $1/\xi$. Furthermore, for all $u \in [0, +\infty)$ it holds that $\mathbb{1}_{\{u \geq 2\xi\}} \leq \Psi_\xi(u) \leq \mathbb{1}_{\{u \geq \xi\}}$. Combining this monotonicity relation with Gine-Zinn symmetrization (see, e.g., [58, Lemma 2.3.1]) and the Rademacher comparison principle for Lipschitz continuous functions (see, e.g., [59, Equa-

tion (4.20)]), we obtain that

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\sum_{\ell=1}^{L}\left(\mathbb{E}\left[\mathbb{1}_{\{f(X_\ell)\geq 2\xi\}}\right]-\mathbb{1}_{\{f(X_\ell)\geq\xi\}}\right)\right]$$

$$\leq\mathbb{E}\left[\sup_{f\in\mathcal{F}}\sum_{\ell=1}^{L}\left(\mathbb{E}\left[\Psi_\xi\left(f\left(X_\ell\right)\right)\right]-\Psi_\xi\left(f\left(X_\ell\right)\right)\right)\right]$$

$$\leq 2\,\mathbb{E}\left[\sup_{f\in\mathcal{F}}\sum_{\ell=1}^{L}\varepsilon_\ell\Psi_\xi\left(f\left(X_\ell\right)\right)\right]$$

$$\leq\frac{2}{\xi}\,\mathbb{E}\left[\sup_{f\in\mathcal{F}}\sum_{\ell=1}^{L}\varepsilon_\ell f\left(X_\ell\right)\right].$$

Together with the inequality chains (47) and (48), this completes the proof. □

A.2. Proof of Lemma A.2

First, we observe that

$$\mathbb{E}\left[\sup_{X\in\mathcal{E}}\operatorname{Re}\left(\sum_{\ell=1}^{L}b_\ell^* X c_\ell\right)\right]=\mathbb{E}\left[\sup_{X\in\mathcal{E}}\operatorname{Re}\left(\langle X,\sum_{\ell=1}^{L}b_\ell c_\ell^*\rangle_F\right)\right]$$

Note that due to the definition of $\omega\left(\mathcal{E}\right)$ in order to finish the proof it is enough to show that the entries of the matrix $X=\sum_{\ell=1}^{L}b_\ell c_\ell^*$ are independent and identically distributed with distribution $\mathcal{CN}\left(0,1\right)$. For that, let $(i,j)\in[K]\times[N]$ and compute that

$$\mathbb{E}\left[|e_i^*\left(\sum_{\ell=1}^{L}b_\ell c_\ell^*\right)e_j|^2\right]=\sum_{\ell=1}^{L}e_i^* b_\ell b_\ell^* e_i\mathbb{E}\left[|c_\ell^* e_j|^2\right]^2=\sum_{\ell=1}^{L}e_i^* b_\ell b_\ell^* e_i=1.$$

This implies that $e_i^*\left(\sum_{\ell=1}^{L}b_\ell c_\ell^*\right)e_j\in\mathcal{CN}\left(0,1\right)$. It remains to show that the individual entries of the matrix $\sum_{\ell=1}^{m}b_\ell c_\ell^*$ are independent. For that, we set

$$X_{i,j}:=\left(\sum_{\ell=1}^{m}b_\ell c_\ell^*\right)_{i,j}=e_i^*\left(\sum_{\ell=1}^{L}b_\ell c_\ell^*\right)e_j.$$

Now let $(i,j),(i',j')\in[K]\times[N]$ such that $(i,j)\neq(i',j')$. Our goal is to show that $\mathbb{E}\left[X_{i,j}\overline{X}_{i',j'}\right]=0$. If $j\neq j'$ this follows immediately from the observation that $c_\ell^* e_j$ and $c_\ell^* e_{j'}$ are independent for all $\ell\in[L]$. Now assume that $j=j'$. Then we can compute that

$$\mathbb{E}\left[X_{i,j}\overline{X}_{i',j'}\right]=\sum_{\ell=1}^{L}e_i^* b_\ell b_\ell^* e_{i'}|c_\ell^* e_j|^2=\sum_{\ell=1}^{L}e_i^* b_\ell b_\ell^* e_{i'}=0.$$

Hence, we have shown that all entries of the matrix X are uncorrelated. As the entries of X are jointly Gaussian this implies that they are independent, which completes the proof.

□

B. Proof of Lemma 5.3

Proof of Lemma 5.3. Let $\ell \in [L]$ such that $\|X^* b_\ell\| \geq 4\xi$. Using the Paley-Zygmund inequality (see, e.g., [60]) we obtain that

$$
\begin{aligned}
\mathbb{P}\left(|\langle b_\ell c_\ell^*, X\rangle_F| \geq 2\xi\right) &\geq \mathbb{P}\left(|\langle b_\ell c_\ell^*, X\rangle_F| \geq \frac{1}{2}\|X^* b_\ell\|\right) \\
&\geq \frac{\left(\mathbb{E}\left[|\langle b_\ell c_\ell^*, X\rangle_F|^2\right] - \frac{1}{4}\|X^* b_\ell\|^2\right)^2}{\mathbb{E}\left[|\langle b_\ell c_\ell^*, X\rangle_F|^4\right]} \\
&= \frac{\left(\|X^* b_\ell\|^2 - \frac{1}{4}\|X^* b_\ell\|^2\right)^2}{2\|X^* b_\ell\|^4} = \frac{9}{32}.
\end{aligned}
$$

(We used that $\mathbb{E}|\langle b_\ell c_\ell^*, X\rangle_F|^2 = \|X^* b_\ell\|^2$ and $\mathbb{E}|\langle b_\ell c_\ell^*, X\rangle_F|^4 = 2\|X^* b_\ell\|^4$, which is due to $\langle b_\ell c_\ell^*, X\rangle_F \sim \mathcal{CN}\left(0, \|X^* b_\ell\|\right)$.) Summing over all $\ell \in [L]$ such that $\|X^* b_\ell\| \geq 4\xi$ yields the claim. $\square$