A convex approach to blind deconvolution with diverse inputs

Ali Ahmed
MIT Mathematics
Email: alikhan@mit.edu

Augustin Cosse
Harvard IACS, UCL ICTEAM
Email: acosse@seas.harvard.edu

Laurent Demanet
MIT Mathematics
Email: laurent@math.mit.edu

***Abstract*—This note considers the problem of blind identification of a linear, time-invariant (LTI) system when the input signals are unknown, but belong to sufficiently diverse, known subspaces. This problem can be recast as the recovery of a rank-1 matrix, and is effectively relaxed using a semidefinite program (SDP). We show that exact recovery of both the unknown impulse response, and the unknown inputs, occurs when the following conditions are met: (1) the impulse response function is spread in the Fourier domain, and (2) the N input vectors belong to generic, known subspaces of dimension K in $\mathbb{R}^L$. Recent results in the well-understood area of low-rank recovery from underdetermined linear measurements can be adapted to show that exact recovery occurs with high probablility (on the genericity of the subspaces) provided that $K, L,$ and N obey the information-theoretic scalings, namely $L \gtrsim K$ and $N \gtrsim 1$ up to log factors.**

I. Introduction

We observe the circular convolutions $\boldsymbol{y}_n = \boldsymbol{h} * \boldsymbol{x}_n$ of N inputs $\{\boldsymbol{x}_n\}_n$ with the system impulse response $\boldsymbol{h}$. Assume that each input $\boldsymbol{x}_n$ respectively lives in a known K-dimensional subspace of $\mathbb{R}^L$, i.e.,

$$\boldsymbol{x}_n = \boldsymbol{C}_n \boldsymbol{m}_n, \ n \in [N],$$

for some $L \times K$ basis matrices $\{\boldsymbol{C}_n\}_{n\in[N]}$. Given the basis matrices, all we need to know are the expansion coefficients $\boldsymbol{m}_n$ to discover the inputs $\boldsymbol{x}_n$ for each n. To express diversity, or genericity, we assume that the elements of $\boldsymbol{C}_n$ are independent $\mathcal{N}(0,1)$ random variables.

This model applies, for instance, in blind channel estimation using random codes. A sequence of length-K messages $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_N$ are respectively coded using tall $L \times K$ coding matrices $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_N$. The protected coded messages $\boldsymbol{x}_n = \boldsymbol{C}_n \boldsymbol{m}_n$ are then transmitted one after the other over a channel with unknown impulse response $\boldsymbol{h} \in \mathbb{R}^L$. The receiver obtains overlapped copies with different delays of each of the N transmitted signals, which can be modeled as the convolutions of the coded messages with the channel response. We are assuming here that the channel's impulse response is fixed over the duration of the transmission of these N coded messages, which justifies the use of a same $\boldsymbol{h}$ in each of the convolutions. Assuming the knowledge of the coding matrices, the task of the decoder is to discover both the channel response and the messages.

The solution proposed in this paper follows a long line of recent work on convex relaxation of nonlinear problems, including matrix completion and phase retrieval. See for example [1], [2], [3], [4], [5], [6], [7]. The idea is to lift the original problem into a problem with linear constraints in a higher-dimensional space, and relax any rank condition with a convex substitute.

To follow this prescription for the problem at hand, let $\boldsymbol{m}^{\mathrm{H}} = [\boldsymbol{m}_1^{\mathrm{H}}, \ldots, \boldsymbol{m}_N^{\mathrm{H}}]$, and write the measurements as

$$\boldsymbol{y}_n(\ell) = \langle \boldsymbol{A}_{\ell,n}, \boldsymbol{h}\boldsymbol{m}^{\mathrm{H}} \rangle_{\mathrm{F}} = \mathrm{tr}(\boldsymbol{A}_{\ell,n} \boldsymbol{m} \boldsymbol{h}^{\mathrm{H}}).$$

A short calculation shows that $\boldsymbol{A}_{\ell,n} = \boldsymbol{f}_\ell \boldsymbol{\phi}_{\ell,n}^{\mathrm{H}}$, where

- $\boldsymbol{f}_\ell^{\mathrm{H}}$ is the ℓth row of the $L \times L$ orthonormal DFT matrix $\boldsymbol{F}$,
- $\boldsymbol{\phi}_{\ell,n}$ is a length KN vector of zeros except for a length K vector $\boldsymbol{c}_{\ell,n}$ placed in the position indexed by $(n-1)K < k \leq nK$ (or $k \sim_K n$ for short), and
- $\bar{\boldsymbol{c}}_{\ell,n} = \sqrt{L} \boldsymbol{C}_n^{\mathrm{H}} \boldsymbol{f}_\ell$.

In the sequel, we also write $\mathcal{A}(\boldsymbol{X})$ for the collection $\langle \boldsymbol{A}_{\ell,n}, \boldsymbol{X} \rangle_{\mathrm{F}}$ indexed by ℓ and n.

The problem of recovering $\boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}$ from $\boldsymbol{y}_n$ is thus equivalently written as

$$\begin{aligned} \text{minimize} \quad & \mathrm{rank}(\boldsymbol{X}) \\ \text{subject to} \quad & \{\boldsymbol{y}_n\} = \mathcal{A}(\boldsymbol{X}). \end{aligned} \tag{1}$$

The unique solution of this program is $\boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}$, which has rank 1. From the knowledge of $\boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}$, observe that $\boldsymbol{h}$ and $\boldsymbol{m}^{\mathrm{H}}$ can only be known up to an inconsequential multiplicative scalar, i.e., the singular vectors of $\boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}$ may be $\alpha \boldsymbol{h}$ and $\boldsymbol{m}^{\mathrm{H}}/\alpha$ for some $\alpha \neq 0$.

Since rank minimization under linear constraints is NP hard in general, we consider the usual relaxation

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{X}\|_* \\ \text{subject to} \quad & \{\boldsymbol{y}_n\} = \mathcal{A}(\boldsymbol{X}), \end{aligned} \tag{2}$$

where $\|\boldsymbol{X}\|_*$ denotes the nuclear norm of $\boldsymbol{X}$ (the sum of its singular values.) We arrive at a convex program; section IV adapts a well-known algorithm from semidefinite programming for solving it in a scalable way.

Success of recovery of $\boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}$ as the unique minimizer of (2) hinges on the following two quantities (coherences) being small:

$$\mu_h^2 := L \cdot \frac{\|\boldsymbol{F}\boldsymbol{h}\|_\infty^2}{\|\boldsymbol{h}\|_2^2}, \qquad \mu_m^2 := N \cdot \max_n \frac{\|\boldsymbol{m}_n\|_2^2}{\|\boldsymbol{m}\|_2^2}. \tag{3}$$

Note that $1 \leq \mu_h^2 \leq L$ and $1 \leq \mu_m^2 \leq N$. The impulse response coherence μ_h is small when $\boldsymbol{h}$ is for instance a sparse train of Diracs. This example is particularly relevant: it corresponds to the case when $\boldsymbol{h}$ realizes a small number of unknown delays (multipath arrivals) in the blind channel estimation application. The message coherence μ_m is small when the N message signals have comparable transmitter energy.

Notice that the framework in this paper differs from that in [5] in that the impulse response $\boldsymbol{h}$ is arbitrary, rather than a member of a known subspace. It is the "diversity" afforded by the presence of a small number of random inputs, in contrast to a single random input, that enables us to consider an arbitrary $\boldsymbol{h}$.

II. Main Result

Our main result states that recovery of both $\boldsymbol{h}$ and $\boldsymbol{m}^{\mathrm{H}}$ is possible as soon as the matrices $\boldsymbol{C}_n$ are taller than wide, and for a number of inputs that depends very weakly on the length L of each signal. The scalings degrade gracefully as the coherence parameters grow.

Theorem 1. *There exist constants for the $\mathcal{O}(\cdot)$ and $\gtrsim$ notations, such that for any $\beta > 0$, if*

$$L \gtrsim \beta \mu_h^2 K \log^6 L, \tag{4}$$

$$N \gtrsim \beta \mu_m^2 \log^6 L, \tag{5}$$

then $\boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}$ is the unique solution to (2) *with probability $1 - \mathcal{O}(L^{-\beta})$.*

III. Mathematical Argument

The proof of Theorem 1 relies on the notion of dual certificate. Let us first introduce the subspace T as

$$T = \{\boldsymbol{h}\boldsymbol{x}^{\mathrm{H}} + \boldsymbol{y}\boldsymbol{m}^{\mathrm{H}} |\ \boldsymbol{x} \in \mathbb{R}^{LK},\ \boldsymbol{y} \in \mathbb{R}^L\}, \tag{6}$$

together with the associated projector

$$\mathcal{P}_T(\boldsymbol{X}) = \boldsymbol{h}\boldsymbol{h}^{\mathrm{H}}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{m}\boldsymbol{m}^{\mathrm{H}} - \boldsymbol{h}\boldsymbol{h}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{m}\boldsymbol{m}^{\mathrm{H}}. \tag{7}$$

The projection onto the orthogonal complement of T is given by $\mathcal{P}_{T^\perp} = \boldsymbol{I} - \mathcal{P}_T(\boldsymbol{X})$.

It was established in [5] (see [2], [8] for important background material) that $\boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}$ is the unique minimizer of (2) when there exists a (dual certificate) $\boldsymbol{Y}$ such that

$$\|\mathcal{P}_T(\boldsymbol{Y}) - \boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}\|_F \leq \frac{1}{2\sqrt{2}\gamma}, \tag{8}$$

$$\|\mathcal{P}_{T^\perp}(\boldsymbol{Y})\| < \frac{1}{2}, \tag{9}$$

where γ is any bound on the operator norm $\|\mathcal{A}\|$. To compute the latter, we observe that

$$\|\mathcal{A}\|^2 \leq \max_{\ell,n} \|\boldsymbol{\phi}_{\ell,n}\|_2^2,$$

which is a maximum over LN independent chi-squared random variables with K degrees of freedom. Hence

$$\mathrm{P}\left(\|\mathcal{A}\|^2 > \lambda\right) \leq 1.2LNe^{\frac{-\lambda}{8K}}$$

In particular, taking $\lambda = \mathcal{O}(K(\log(LN) + \beta \log L))$ we have $\|\mathcal{A}\|^2 \lesssim K(\log(LN) + \beta\log(L))$, with probability $1 - L^{-\beta}$. Thus, we let $\gamma = C\sqrt{K(\log(LN) + \beta\log(L))}$ for some adequately large $C > 0$.

To build a certificate $\boldsymbol{Y}$ satisfying equations (8) and (9), we use the golfing scheme introduced in [8]. Let Γ be the set of all (ℓ, n) indexing the measurements $\boldsymbol{Y}_{\ell,n}$. For $1 \leq P \leq LN$, we introduce the partition $\{\Gamma_p\}_{p=1}^P$ obtained by randomly choosing elements from Γ without replacement and grouping them into sets $\Gamma_1, \ldots, \Gamma_P$ of equal cardinality. (The condition that $\bigcup_p \Gamma_p$ covers Γ can be relaxed when P is not a divisor of LN.) The particular value of P that makes the argument below work will be determined later.

The golfing scheme then iteratively defines the certificate as

$$\boldsymbol{Y}_{p+1} = \boldsymbol{Y}_p + \frac{|\Gamma|}{|\Gamma_p|}\mathcal{A}_p^{\mathrm{H}}\mathcal{A}_p(\boldsymbol{h}\boldsymbol{m}^{\mathrm{H}} - \mathcal{P}_T(\boldsymbol{Y}_p)), \tag{10}$$

where $\mathcal{A}_p$ is the restriction of $\mathcal{A}$ to indices in Γ_p.

In order to show that $\boldsymbol{Y}_P$ obtained by golfing satisfies the conditions (8) and (9), we need two auxiliary lemmas. Their proofs will appear in the long version of this note, and are inspired by similar developments in [5].

Lemma 1. *Consider the partitioning $\{\Gamma_p\}_p$ as above. Then, with probability $1 - \mathcal{O}(L^{-\beta})$,*

$$\max_{p=1,\ldots,P} \|\frac{|\Gamma|}{|\Gamma_p|}\sum_{(\ell,n)\in\Gamma_p} \mathcal{P}_T\mathcal{A}_{\ell,n}^{\mathrm{H}}\mathcal{A}_{\ell,n}\mathcal{P}_T - \mathcal{P}_T\| \leq \frac{1}{2},$$

provided $L \gtrsim \beta\mu_h^2 K \log^3 L$, $N \gtrsim \beta\mu_m^2 \log^3 L$.

To state the second lemma, we first need to introduce the following extensions to (3). For any matrix $\boldsymbol{B} \in \mathbb{C}^{L\times NK}$ with column wise Fourier transform denoted by $\hat{\boldsymbol{B}} = \boldsymbol{F}\boldsymbol{B}$,

$$\begin{aligned} \mu_h^2(\boldsymbol{B}) &= \frac{L}{\|\boldsymbol{B}\|_{\mathrm{F}}^2} \max_{\ell\in[L]} \sum_{n\in[N]} \sum_{k\sim_K n} \hat{B}^2[\ell,k], \\ \mu_m^2(\boldsymbol{B}) &= \frac{N}{\|\boldsymbol{B}\|_{\mathrm{F}}^2} \max_{n\in[L]} \sum_{\ell\in[L]} \sum_{k\sim_K n} \hat{B}^2[\ell,k]. \end{aligned} \tag{11}$$

We consider the special case when $\boldsymbol{B} = \boldsymbol{W}_p$, with

$$\boldsymbol{W}_p = \mathcal{P}_T(\boldsymbol{Y}_p) - \boldsymbol{h}\boldsymbol{m}^{\mathrm{H}},$$

so that

$$\mu_{h_p}^2 := \mu_h^2(\boldsymbol{W}_p), \qquad \mu_{m_p}^2 := \mu_m^2(\boldsymbol{W}_p). \tag{12}$$

It is possible to show (see [5], lemma 4) that $\mu_{h_p}^2 \leq \mu_h^2$ and $\mu_{m_p}^2 \leq \mu_m^2$ with high probability.

Lemma 2. *Consider the partitioning $\{\Gamma_p\}_p$ as above. Let $\boldsymbol{B}$ be deterministic, or random independent from $\boldsymbol{A}_{\ell,n}$ for $(\ell,n) \in \Gamma_p$. Then for $\beta > 1$, with probability at least $1 - \mathcal{O}(L^{-\beta})$,*

$$\|\frac{|\Gamma|}{|\Gamma_p|}\sum_{(\ell,n)\in\Gamma_p} \langle \boldsymbol{B}, \boldsymbol{A}_{\ell,n}\rangle \boldsymbol{A}_{\ell,n} - \boldsymbol{B}\| \leq \frac{1}{2^{p+1}},$$

when $L \gtrsim \beta\mu_h^2(\boldsymbol{B})K\log^6(L)$ and $N \gtrsim \beta\mu_m^2(\boldsymbol{B})\log^6(L)$.

We apply this lemma to $\boldsymbol{B} = \boldsymbol{W}_{p-1}$, which is manifestly independent from $\boldsymbol{A}_{\ell,n}$ for $(\ell, n) \in \Gamma_p$ since the subsets $\{\Gamma_p\}_p$ are disjoint. Hence we obtain

$$\|\frac{|\Gamma|}{|\Gamma_p|} \sum_{(\ell,n)\in\Gamma_p} \langle \boldsymbol{W}_{p-1}, \boldsymbol{A}_{\ell,n}\rangle \boldsymbol{A}_{\ell,n} - \boldsymbol{W}_{p-1}\| \leq \frac{1}{2^{p+1}} \quad (13)$$

with high probability, under the same conditions on L and N as in theorem 1.

The proofs of lemma 1 and lemma 2 (not given here) involve a careful treatment of different variants of sampling for Γ_p, such as Bernoulli sampling and uniform sampling with replacement. The proofs hinge on the fact that these variants are more adequate in some contexts, yet provide a sufficiently similar degree of randomness to uniform sampling without replacement.

Using lemmas 1 and 2, we now show that the certificate given by golfing in (10) satisfies the conditions (8) and (9) required to prove the recovery.

First, using the recursion (10), we relate $\boldsymbol{W}_p$ to $\boldsymbol{W}_{p-1}$ via

$$\begin{aligned}\boldsymbol{W}_p =& \mathcal{P}_T(\boldsymbol{Y}_p) - \boldsymbol{h}\boldsymbol{m}^{\mathrm{H}} \\ =& \mathcal{P}_T(\boldsymbol{Y}_{p-1}) - \frac{|\Gamma|}{|\Gamma_p|}\mathcal{P}_T\mathcal{A}_p^{\mathrm{H}}\mathcal{A}_p\mathcal{P}_T(\mathcal{P}_T(\boldsymbol{Y}_{p-1} - \boldsymbol{h}\boldsymbol{m}^{\mathrm{H}})) \\ & - \boldsymbol{h}\boldsymbol{m}^{\mathrm{H}} \\ =& \left(\mathcal{P}_T - \frac{|\Gamma|}{|\Gamma_p|}\mathcal{P}_T\mathcal{A}_p^{\mathrm{H}}\mathcal{A}_p\mathcal{P}_T\right)\boldsymbol{W}_{p-1}.\end{aligned}$$

By using lemma 1 recursively on this last equality, we get

$$\|\mathcal{P}_T(\boldsymbol{Y}_p) - \boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}\| = \|\boldsymbol{W}_p\| \leq 2^{-p}, \qquad p = 1, \ldots, P. \quad (14)$$

Hence with $P = (|\Gamma|/|\Gamma_p|) \geq \log_2(2\sqrt{2}\gamma)$, we satisfy the first condition, equation (8).

Second, we easily compute that

$$\boldsymbol{Y}_P = -\sum_{p=1}^{P} \frac{|\Gamma|}{|\Gamma_p|}\mathcal{A}_p^{\mathrm{H}}\mathcal{A}_p\boldsymbol{W}_{p-1}. \quad (15)$$

By noting that $\boldsymbol{W}_{p-1}$ belongs to T, (so that $\mathcal{P}_{T^\perp}(\boldsymbol{W}_{p-1}) = 0$,) and because $\|\mathcal{P}_{T^\perp}\boldsymbol{X}\| \leq \|\boldsymbol{X}\|$, we get

$$\|\mathcal{P}_{T^\perp}(\boldsymbol{Y}_p)\| \leq \sum_{p=1}^{P} \frac{|\Gamma|}{|\Gamma_p|}\left\|\mathcal{A}_p^{\mathrm{H}}\mathcal{A}_p\boldsymbol{W}_{p-1} - \frac{|\Gamma_p|}{|\Gamma|}\boldsymbol{W}_{p-1}\right\|. \quad (16)$$

After applying lemma 2, this last inequality can be shown to satisfy the condition of equation (9), namely $\|\mathcal{P}_{T^\perp}(\boldsymbol{Y}_p)\| < \frac{1}{2}$.

IV. Numerical simulations

In this section we conduct numerical experiments and study the resulting phase diagrams in order to quantify the probability of success for the formulation (2).

To conduct the numerical experiments, we first write problem (2) as a trace minimization problem (see [9]),

$$\begin{aligned} \min \quad & \mathrm{Tr}(\boldsymbol{V}_{11}) + \mathrm{Tr}(\boldsymbol{V}_{22}) \\ s.t \quad & \mathcal{A}(\boldsymbol{V}_{12}) = \{\boldsymbol{y}_n\}, \quad \boldsymbol{V} \succeq 0. \end{aligned} \quad (17)$$

Here $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix}$ is introduced as a proxy for the rank-1 matrix $\boldsymbol{V}_0 = \begin{pmatrix} \boldsymbol{h}\boldsymbol{h}^H & \boldsymbol{h}\boldsymbol{m}^H \\ \boldsymbol{m}\boldsymbol{h}^H & \boldsymbol{m}\boldsymbol{m}^H \end{pmatrix}$. To handle reasonably large instances of (17), we follow the approach in [10] and consider a low rank factorization for the matrix $\boldsymbol{V}$ as $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{R}_1 \\ \boldsymbol{R}_2 \end{pmatrix}\begin{pmatrix} \boldsymbol{R}_1^H & \boldsymbol{R}_2^H \end{pmatrix}$, where $\boldsymbol{R}_1 \in \mathbb{C}^{L\times r}$, $\boldsymbol{R}_2 \in \mathbb{C}^{KN\times r}$ for some rank $r \ll L + KN$. We fix $r = 4$ for convenience. We then minimize the augmented Lagrangian $\mathcal{L}(\boldsymbol{R})$ associated to (17) with respect to $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 \\ \boldsymbol{R}_2 \end{bmatrix}$, where

$$\begin{aligned}\mathcal{L}(\boldsymbol{R}) =& \frac{1}{2}\|\boldsymbol{R}_1\|_F^2 + \frac{1}{2}\|\boldsymbol{R}_2\|_F^2 + \sum_{\ell,n}(\langle \boldsymbol{A}_{\ell,n}, \boldsymbol{R}_1\boldsymbol{R}_2^H\rangle - \boldsymbol{y}_n(\ell)) \\ & + \sum_{\ell,n}\frac{\sigma}{2}(\langle \boldsymbol{A}_{\ell,n}, \boldsymbol{R}_1\boldsymbol{R}_2^H\rangle - \boldsymbol{y}_n(\ell))^2. \end{aligned} \quad (18)$$

The success rates for different values of the parameters L, N and K are shown in Fig. 1. The first set of numerical experiments shows the rate of success as a function of both the size of the input subspace K and the size of the data space L, N being fixed to 40. For each of the values of the pairs (K, L), 100 experiments were run by taking gaussian i.i.d sensing matrices $\boldsymbol{C}_n$ with $\boldsymbol{C}_n(\ell, k) \sim \mathcal{N}(0, 1)$, as well as gaussian i.i.d vectors $\boldsymbol{h}$ and $\boldsymbol{m}$. For each of the 100 experiments, we ran 40 iterations of LBFGS on the augmented Lagrangian (18). An experiment was classified as a success (1) as opposed to failure (0) whenever the relative difference between the obtained matrix $\boldsymbol{X}$ and the optimal solution $\boldsymbol{X}_0 = \boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}$ was less than 2%. In other words,

$$\frac{\|\boldsymbol{X} - \boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}\|_F}{\|\boldsymbol{h}\boldsymbol{m}^{\mathrm{H}}\|_F} < .02 \quad (19)$$

The second diagram of Figure 1 shows the rate of success as a function of K and N for the same experimental framework. Here L is fixed to 800.

Finally the third diagram shows the rate of success for various values of L and N for a fixed $K = 40$.

V. Conclusion

In this note, we consider a version of the blind deconvolution problem where the tradeoff is moved towards an arbitrary transfer function, at the expense of requiring a small number of sufficiently different input signals to guarantee the recovery. Theory shows that whenever the dimensions of the problem N, L and K satisfy $L \gtrsim \beta K\mu_h^2 \log^6 L$ and $N \gtrsim \beta\mu_m^2 \log^6 L$, both the impulse response function as well as the unknown signals can be recovered up to a scaling factor from program (2). Finally, together with a computationally tractable implementation for (2), we provide numerical evidence of the recovery by means of phase diagrams.

Acknowledgment

AA is supported by ONR and NSF. AC is supported by the FNRS and MISTI Belgium. LD is supported by AFOSR, ONR, NSF, and Total SA.

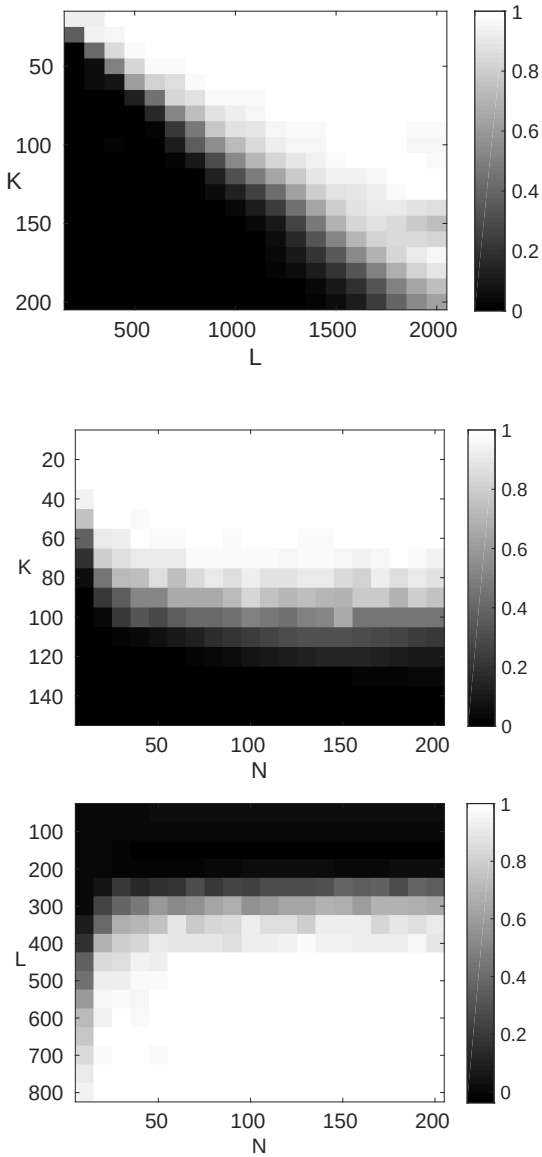

Fig. 1. Recovery of $\boldsymbol{h}$ and $\boldsymbol{m}$ from nuclear norm minimization for (top) various values of K and L, (middle) various values of K and N, and (bottom) various values of L and N. White indicates success while black indicates failure. From the top figure, the empirical recovery rate is seen to increase when L increases and when K decreases, with an expected phase transition when L is a multiple of K. From the middle and bottom figures, we see that a minimum (small) value of N is needed for recovery, but past this threshold, K and L are the only parameters affecting recovery. All these phase transitions can be explained by noting that the number of unknowns in the original problem is given by $KN + L$ whether the number of measurements is given by LN.

REFERENCES

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

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

[3] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[4] A. Ahmed and J. Romberg, "Compressive multiplexing of correlated signals," *arXiv preprint arXiv:1308.5146*, 2013.

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

[6] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Review*, vol. 57, no. 2, pp. 225–251, 2015.

[7] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *arXiv preprint arXiv:1501.06864*, 2015.

[8] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *Information Theory, IEEE Transactions on*, vol. 57, no. 3, pp. 1548–1566, 2011.

[9] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *American Control Conference, 2001. Proceedings of the 2001*, vol. 6. IEEE, 2001, pp. 4734–4739.

[10] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Mathematical Programming*, vol. 95, no. 2, pp. 329–357, 2003.