

Supplementary Note:

Blind Deconvolutional Phase Retrieval via Convex Programming

Abstract

The material presented in this document is supplementary to the manuscript submitted to NIPS 2018. The document contains extended discussions, additional technical proofs and details of the convex program implementation.

1 Visible Light Communication

As discussed in the body of the paper, an important application domain where blind deconvolution from phaseless Fourier measurements arises is the visible light communication (VLC). A stylized VLC setup is shown in Figure 1. A message $\mathbf{m} \in \mathbb{R}^n$ is to be transmitted using visible light. The message is first coded by multiplying it with a tall coding matrix $\mathbf{C} \in \mathbb{R}^{m \times n}$ and the resultant information $\mathbf{x} = \mathbf{C}\mathbf{m}$ is modulated on a light wave. The light wave propagates through an unknown media. This propagation can be modeled as a convolution $\mathbf{x} \circledast \mathbf{w}$ of the information signal \mathbf{x} with unknown channel $\mathbf{w} \in \mathbb{R}^m$. The vector \mathbf{w} contains channel taps, and frequently in realistic applications has only few significant taps. In this case, one can model

$$\mathbf{w} \approx \mathbf{B}\mathbf{h},$$

where $\mathbf{h} \in \mathbb{R}^k$ is a short ($k \ll m$) vector, and $\mathbf{B} \in \mathbb{R}^{m \times k}$ in this case is a subset of the columns of an identity matrix. Generally, the multipath channels are well modeled with non-zero taps in top locations of \mathbf{w} . In that case, \mathbf{B} is exactly known to be top few columns of the identity matrix.

In visible light communication, there is always a difficulty associated with measuring phase information in the received light. Figure 1 shows a setup, where we measure the phaseless Fourier transform (light through the lens) of this signal. The measurements are therefore

$$\mathbf{y} = |\mathbf{F}(\mathbf{C}\mathbf{m} \circledast \mathbf{B}\mathbf{h})|$$

and one wants to recover \mathbf{m} , and \mathbf{h} given the knowledge of \mathbf{B} , and the coding matrix \mathbf{C} . Since we chose \mathbf{C} to be random Gaussian, and \mathbf{B} is the columns of identity. As mentioned at the

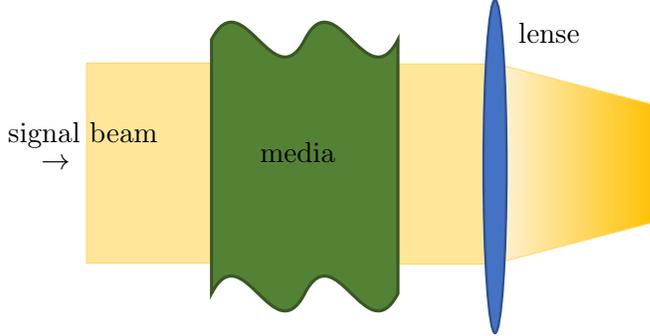


Figure 1: Visible light communication optical setup; the media block normally consists of phosphor, filter and a linear polarizer. The lens takes the Fourier transform of the light and one can only measure the intensity only measurements of this transformed light source signal.

end of the numerics section that with this subspace model, we obtain similar recovery results as one would have for both \mathbf{B} , and \mathbf{C} being random Gaussians. The proposed convex program solves this difficult inverse problem and recovers the true solution with these subspace models.

2 Proof of Lemma 3

The proof is based on small ball method developed in [Koltchinskii and Mendelson(2015), Mendelson(2014)] and further studied in [Lecué et al.(2018)Lecué, Mendelson, et al.] and [Lecué and Mendelson(2017)]. The proof is mainly repeated using a similar line of argument as in [Bahmani and Romberg(2017)], and is provided here for completeness.

Rest of the proof now concerns showing that $(\tilde{\mathbf{H}}, \tilde{\mathbf{M}})$ is the unique solution to the linearly constrained optimization program (9). Define one sided loss function:

$$\mathcal{L}(\mathbf{H}, \mathbf{M}) := \sum_{\ell=1}^m \left(2y_{\ell}^2 - \frac{1}{m} \langle \mathbf{b}_{\ell} \mathbf{b}_{\ell}^*, \mathbf{H} \rangle \langle \mathbf{c}_{\ell} \mathbf{c}_{\ell}^*, \tilde{\mathbf{M}} \rangle - \langle \mathbf{b}_{\ell} \mathbf{b}_{\ell}^*, \tilde{\mathbf{H}} \rangle \langle \mathbf{c}_{\ell} \mathbf{c}_{\ell}^*, \mathbf{M} \rangle \right)_+, \quad (\text{S.1})$$

where $(\cdot)_+$ denotes the positive side. Using this definition, we rewrite (9) compactly as

$$\begin{aligned} & \text{minimize } \|\mathbf{H}\|_* + \|\mathbf{M}\|_* & (\text{S.2}) \\ & \text{subject to } \mathcal{L}(\mathbf{H}, \mathbf{M}) \leq 0. \end{aligned}$$

The goal of the proof is to show that all descent direction $(\delta\mathbf{H}, \delta\mathbf{M}) \in \mathcal{Q}$ that also obey the constraint set have a small ℓ_2 norm. Since $(\delta\mathbf{H}, \delta\mathbf{M})$ is a feasible perturbation from the proposed optimal $(\tilde{\mathbf{H}}, \tilde{\mathbf{M}})$, we have from the constraints above that

$$\mathcal{L}(\tilde{\mathbf{H}} + \delta\mathbf{H}, \tilde{\mathbf{M}} + \delta\mathbf{M}) \leq 0 \quad (\text{S.3})$$

We begin by expanding the loss function $\text{Loss}(\tilde{\mathbf{H}} + \delta\mathbf{H}, \tilde{\mathbf{M}} + \delta\mathbf{M})$ below

$$\begin{aligned}
& \mathcal{L}(\tilde{\mathbf{H}} + \delta\mathbf{H}, \tilde{\mathbf{M}} + \delta\mathbf{M}) \\
&= \sum_{\ell=1}^m \left(2y_\ell^2 - \frac{1}{m} (\langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \tilde{\mathbf{H}} + \delta\mathbf{H} \rangle \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \tilde{\mathbf{M}} \rangle + \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \tilde{\mathbf{H}} \rangle \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \tilde{\mathbf{M}} + \delta\mathbf{M} \rangle) \right)_+ \\
&= \frac{1}{m} \sum_{\ell=1}^m \left((\langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \tilde{\mathbf{H}} \rangle \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \tilde{\mathbf{M}} \rangle + \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \tilde{\mathbf{H}} + \delta\mathbf{H} \rangle \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \tilde{\mathbf{M}} \rangle) - \right. \\
&\quad \left. (\langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \tilde{\mathbf{H}} + \delta\mathbf{H} \rangle \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \tilde{\mathbf{M}} \rangle + \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \tilde{\mathbf{H}} \rangle \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \tilde{\mathbf{M}} + \delta\mathbf{M} \rangle) \right)_+ \\
&= \frac{1}{m} \sum_{\ell=1}^m \left(- \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \delta\mathbf{H} \rangle \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \tilde{\mathbf{M}} \rangle - \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \tilde{\mathbf{H}} \rangle \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \delta\mathbf{M} \rangle \right)_+ \\
&\geq \frac{1}{m} \sum_{\ell=1}^m ((-\langle \nabla f_\ell, (\delta\mathbf{H}, \delta\mathbf{M}) \rangle))_+. \tag{S.4}
\end{aligned}$$

where the last equality follows from the using notation $\nabla f_\ell = (\tilde{v}_\ell \mathbf{b}_\ell \mathbf{b}_\ell^*, \tilde{u}_\ell \mathbf{c}_\ell \mathbf{c}_\ell^*)$ introduced earlier. Let $\psi_t(s) := (s)_+ - (s - t)_+$. Using the fact that $\psi_t(s) \leq (s)_+$, and that for every $\alpha, t \geq 0$, and $s \in \mathbb{R}$, $\psi_{\alpha t}(s) = t\psi_\alpha(\frac{s}{t})$, we have

$$\begin{aligned}
& \frac{1}{m} \sum_{\ell=1}^m [- \langle \nabla f_\ell, (\delta\mathbf{H}, \delta\mathbf{M}) \rangle]_+ \geq \frac{1}{m} \sum_{\ell=1}^m \psi_\tau \|(\delta\mathbf{H}, \delta\mathbf{M})\|_F \left(- \langle \nabla f_\ell, (\delta\mathbf{H}, \delta\mathbf{M}) \rangle \right) \\
&= \|(\delta\mathbf{H}, \delta\mathbf{M})\|_F \cdot \frac{1}{m} \sum_{\ell=1}^m \psi_\tau \left(- \langle \nabla f_\ell, \frac{(\delta\mathbf{H}, \delta\mathbf{M})}{\|(\delta\mathbf{H}, \delta\mathbf{M})\|_F} \rangle \right) \\
&= \|(\delta\mathbf{H}, \delta\mathbf{M})\|_F \left[\frac{1}{m} \sum_{\ell=1}^m \mathbb{E} \psi_\tau \left(- \langle \nabla f_\ell, \frac{(\delta\mathbf{H}, \delta\mathbf{M})}{\|(\delta\mathbf{H}, \delta\mathbf{M})\|_F} \rangle \right) \right. \\
&\quad \left. - \frac{1}{m} \sum_{\ell=1}^m \left[\mathbb{E} \psi_\tau \left(- \langle \nabla f_\ell, \frac{(\delta\mathbf{H}, \delta\mathbf{M})}{\|(\delta\mathbf{H}, \delta\mathbf{M})\|_F} \rangle \right) - \psi_\tau \left(- \langle \nabla f_\ell, \frac{(\delta\mathbf{H}, \delta\mathbf{M})}{\|(\delta\mathbf{H}, \delta\mathbf{M})\|_F} \rangle \right) \right] \right]. \tag{S.5}
\end{aligned}$$

Define a centered random process $\mathcal{R}(\mathbf{B}, \mathbf{C})$ as follows

$$\mathcal{R}(\mathbf{B}, \mathbf{C}) := \sup_{(\delta\mathbf{H}, \delta\mathbf{M}) \in \mathcal{Q}} \frac{1}{m} \sum_{\ell=1}^m \left[\mathbb{E} \psi_\tau \left(- \langle \nabla f_\ell, \frac{(\delta\mathbf{H}, \delta\mathbf{M})}{\|(\delta\mathbf{H}, \delta\mathbf{M})\|_F} \rangle \right) - \psi_\tau \left(- \langle \nabla f_\ell, \frac{(\delta\mathbf{H}, \delta\mathbf{M})}{\|(\delta\mathbf{H}, \delta\mathbf{M})\|_F} \rangle \right) \right]$$

and an application of bounded difference inequality [McDiarmid(1989)] yields that $\mathcal{R}(\mathbf{B}, \mathbf{C}) \leq \mathbb{E} \mathcal{R}(\mathbf{B}, \mathbf{C}) + t\tau/\sqrt{m}$ with probability at least $1 - e^{-2mt^2}$. It remains to evaluate $\mathbb{E} \mathcal{R}(\mathbf{B}, \mathbf{C})$, which after using a simple symmetrization inequality [van der Vaart and Wellner(1997)] yields

$$\mathbb{E} \mathcal{R}(\mathbf{B}, \mathbf{C}) \leq 2\mathbb{E} \sup_{(\delta\mathbf{H}, \delta\mathbf{M}) \in \mathcal{Q}} \frac{1}{m} \sum_{\ell=1}^m \varepsilon_\ell \psi_\tau \left(- \langle \nabla f_\ell, \frac{(\delta\mathbf{H}, \delta\mathbf{M})}{\|(\delta\mathbf{H}, \delta\mathbf{M})\|_F} \rangle \right), \tag{S.6}$$

where $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ are independent Rademacher random variables. Using the fact that $\psi_t(0) = 0$, and $\psi_t(s)$ is a contraction: $|\psi_t(\alpha_1) - \psi_t(\alpha_2)| \leq |\alpha_1 - \alpha_2|$ for all $\alpha_1, \alpha_2 \in \mathbb{R}$, we have from the Rademacher contraction inequality Theorem 4.12 in [Ledoux and Talagrand(2013)] that

$$\begin{aligned} \mathbb{E} \sup_{(\delta \mathbf{H}, \delta \mathbf{M}) \in \mathcal{Q}} \frac{1}{m} \sum_{\ell=1}^m \varepsilon_\ell \psi_\tau \left(- \langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \right) &\leq \mathbb{E} \sup_{(\delta \mathbf{H}, \delta \mathbf{M}) \in \mathcal{Q}} \frac{1}{m} \sum_{\ell=1}^m -\varepsilon_\ell \langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \\ &= \mathbb{E} \sup_{(\delta \mathbf{H}, \delta \mathbf{M}) \in \mathcal{Q}} \frac{1}{m} \sum_{\ell=1}^m \varepsilon_\ell \langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle, \end{aligned} \quad (\text{S.7})$$

where the last equality is the result of the fact that a global sign change of a sequence of Rademacher random variables does not change their distribution. In addition, using the facts that $t\mathbf{1}(s \geq t) \leq \psi_t(s)$, and that random vectors $\nabla f_1, \nabla f_2, \dots, \nabla f_m$ are identically distributed and the distribution is symmetric, it follows

$$\begin{aligned} \tau \mathbb{P} \left(\langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \geq \tau \right) &= \tau \mathbb{E} \left(\mathbf{1} \left[\langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \geq \tau \right] \right) \\ &\leq \mathbb{E} \psi_\tau \left(\langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \right). \end{aligned} \quad (\text{S.8})$$

Plugging (S.8), and (S.7) in (S.5), we have

$$\begin{aligned} \frac{1}{m} \sum_{\ell=1}^m \left(- \langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \right)_+ &\geq \tau \|(\delta \mathbf{H}, \delta \mathbf{M})\|_F \cdot \mathbb{P} \left(\langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \geq \tau \right) \\ &- 2 \|(\delta \mathbf{H}, \delta \mathbf{M})\|_F \cdot \mathbb{E} \sup_{(\delta \mathbf{H}, \delta \mathbf{M}) \in \mathcal{Q}} \left[\frac{1}{m} \sum_{\ell=1}^m \varepsilon_\ell \langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \right] - 2 \|(\delta \mathbf{H}, \delta \mathbf{M})\|_F \frac{t\tau}{\sqrt{m}}. \end{aligned}$$

Combining this with (S.3) and (S.4), we obtain the final result

$$\begin{aligned} \|(\delta \mathbf{H}, \delta \mathbf{M})\|_F \left[\tau \mathbb{P} \left(\langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \geq \tau \right) - 2 \mathbb{E} \sup_{(\delta \mathbf{H}, \delta \mathbf{M}) \in \mathcal{Q}} \frac{1}{m} \sum_{\ell=1}^m \varepsilon_\ell \langle \nabla f_\ell, \frac{(\delta \mathbf{H}, \delta \mathbf{M})}{\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F} \rangle \right] \\ - 2 \|(\delta \mathbf{H}, \delta \mathbf{M})\|_F \frac{t\tau}{\sqrt{m}} \leq 0. \end{aligned}$$

Using the definitions in (12), and (13), we can write

$$\|(\delta \mathbf{H}, \delta \mathbf{M})\|_F \left(\tau p_\tau(\mathcal{Q}) - \frac{(2\mathfrak{C}(\mathcal{Q}) + t\tau)}{\sqrt{m}} \right) \leq 0.$$

It is clear that choosing $m \geq \left(\frac{2\mathfrak{C}(\mathcal{Q}) + t\tau}{\tau p_\tau(\mathcal{Q})} \right)^2$ implies

$$(\delta \mathbf{H}, \delta \mathbf{M}) = (\mathbf{0}, \mathbf{0}).$$

The proof is complete.

3 Probability $p_\tau(\mathcal{Q})$

In this section, we determine the probability $p_\tau(\mathcal{Q})$, and the positive parameter τ in (13) for the set \mathcal{Q} in (11). For a point $(\delta\mathbf{H}, \delta\mathbf{M}) \in \mathcal{Q}$, and randomly chosen ∇f_ℓ , we have via Paley Zygmund inequality that

$$\mathbb{P}\left(|\langle \nabla f_\ell, (\delta\mathbf{H}, \delta\mathbf{M}) \rangle|^2 \geq \frac{1}{2} \mathbb{E} |\langle \nabla f_\ell, (\delta\mathbf{H}, \delta\mathbf{M}) \rangle|^2\right) \geq \frac{1}{4} \frac{(\mathbb{E} |\langle \nabla f_\ell, (\delta\mathbf{H}, \delta\mathbf{M}) \rangle|^2)^2}{\mathbb{E} |\langle \nabla f_\ell, (\delta\mathbf{H}, \delta\mathbf{M}) \rangle|^4}.$$

The particular choice of random gradient vectors we are using is $\nabla f_\ell = (\tilde{v}_\ell \mathbf{b}_\ell \mathbf{b}_\ell^*, \tilde{u}_\ell \mathbf{c}_\ell \mathbf{c}_\ell^*)$ giving us $\langle \nabla f_\ell, (\delta\mathbf{H}, \delta\mathbf{M}) \rangle = \tilde{v}_\ell \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \delta\mathbf{H} \rangle + \tilde{u}_\ell \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \delta\mathbf{M} \rangle$. Since \mathbf{b}_ℓ , and \mathbf{c}_ℓ are standard Gaussian vectors, using the equivalence of L_p -norms for Gaussians, we deduce that

$$\left(\mathbb{E} |\tilde{v}_\ell \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \delta\mathbf{H} \rangle + \tilde{u}_\ell \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \delta\mathbf{M} \rangle|^4\right)^{1/4} \leq c \left(\mathbb{E} |\tilde{v}_\ell \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \delta\mathbf{H} \rangle + \tilde{u}_\ell \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \delta\mathbf{M} \rangle|^2\right)^{1/2}.$$

Plugging last two inequalities in (13) reveals that

$$p_\tau(\mathcal{Q}) \geq c > 0 \tag{S.9}$$

for an absolute constant c . To compute τ , we expand $\mathbb{E} |\langle \nabla f_\ell, (\delta\mathbf{H}, \delta\mathbf{M}) \rangle|^2$ giving us

$$\begin{aligned} \mathbb{E} |\tilde{v}_\ell \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \delta\mathbf{H} \rangle + \tilde{u}_\ell \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \delta\mathbf{M} \rangle|^2 &= 3\|\tilde{\mathbf{m}}\|_2^4 (\langle \text{diag}(\delta\mathbf{H}), \delta\mathbf{H} \rangle + 2\|\delta\mathbf{H}\|_F^2) \\ &\quad + 3\|\tilde{\mathbf{h}}\|_2^4 (\langle \text{diag}(\delta\mathbf{M}), \delta\mathbf{M} \rangle + 2\|\delta\mathbf{M}\|_F^2) + 2|\tilde{\mathbf{h}}^* \text{diag}(\delta\mathbf{H}) \tilde{\mathbf{h}} + 2\tilde{\mathbf{h}}^* \delta\mathbf{H} \tilde{\mathbf{h}}|^2, \end{aligned} \tag{S.10}$$

where we have made use of multiple simple facts including that $\mathbb{E} |\tilde{u}_\ell|^2 = 3\|\tilde{\mathbf{h}}\|_2^4$, and similarly for \tilde{v}_ℓ , and two identities: $\mathbb{E} |\mathbf{b}_\ell^* \tilde{\mathbf{h}}|^2 \mathbf{b}_\ell^* \delta\mathbf{H} \mathbf{b}_\ell = \tilde{\mathbf{h}}^* \text{diag}(\delta\mathbf{H}) \tilde{\mathbf{h}} + 2\tilde{\mathbf{h}}^* \delta\mathbf{H} \tilde{\mathbf{h}}$, and $\mathbb{E} (\mathbf{b}_\ell^* \delta\mathbf{H} \mathbf{b}_\ell) \mathbf{b}_\ell \mathbf{b}_\ell^* = \text{diag}(\delta\mathbf{H}) + 2(\delta\mathbf{H}) \implies \mathbb{E} |\mathbf{b}_\ell^* \delta\mathbf{H} \mathbf{b}_\ell|^2 = \langle \text{diag}(\delta\mathbf{H}), \delta\mathbf{H} \rangle + 2\|\delta\mathbf{H}\|_F^2$. We also made use of the fact that $\mathcal{Q} \perp \mathcal{N}$ and therefore $\langle \tilde{\mathbf{H}}, \delta\mathbf{H} \rangle - \langle \tilde{\mathbf{M}}, \delta\mathbf{M} \rangle = 0$, or equivalently, $\tilde{\mathbf{h}}^* \delta\mathbf{H} \tilde{\mathbf{h}} = \tilde{\mathbf{m}}^* \delta\mathbf{M} \tilde{\mathbf{m}}$.

It is easy to conclude from (S.10) now that

$$\begin{aligned} \mathbb{E} |\tilde{v}_\ell \langle \mathbf{b}_\ell \mathbf{b}_\ell^*, \delta\mathbf{H} \rangle + \tilde{u}_\ell \langle \mathbf{c}_\ell \mathbf{c}_\ell^*, \delta\mathbf{M} \rangle|^2 &\geq 6(\|\tilde{\mathbf{h}}\|_2^4 \|\delta\mathbf{H}\|_F^2 + \|\tilde{\mathbf{m}}\|_2^4 \|\delta\mathbf{M}\|_F^2) \\ &\geq c \min(\|\tilde{\mathbf{h}}\|_2^2, \|\tilde{\mathbf{m}}\|_2^2) (\|\delta\mathbf{H}\|_F^2 + \|\delta\mathbf{M}\|_F^2) = c \max(\|\tilde{\mathbf{h}}\|_2^2, \|\tilde{\mathbf{m}}\|_2^2) (\|\delta\mathbf{H}\|_F^2 + \|\delta\mathbf{M}\|_F^2), \end{aligned}$$

where the last equality uses the fact that $\text{Tr}(\tilde{\mathbf{H}}) = \text{Tr}(\tilde{\mathbf{M}})$ from (8), which is equivalent to $\|\tilde{\mathbf{h}}\|_2^2 = \|\tilde{\mathbf{m}}\|_2^2$. This directly means, we can take $\tau = c \max(\|\tilde{\mathbf{h}}\|_2^2, \|\tilde{\mathbf{m}}\|_2^2)$, where c is an absolute constant.

4 Implementing the Convex Program

In this section we take an alternating direction method of multipliers (ADMM) scheme to address (5), which takes the form

$$\begin{aligned}
& \underset{\mathbf{X}_1, \mathbf{X}_2}{\text{minimize}} && \text{Tr}(\mathbf{X}_1) + \text{Tr}(\mathbf{X}_2) && \text{(S.11)} \\
& \text{subject to} && \langle \mathbf{a}_{1,\ell} \mathbf{a}_{1,\ell}^*, \mathbf{X}_1 \rangle \langle \mathbf{a}_{2,\ell} \mathbf{a}_{2,\ell}^*, \mathbf{X}_2 \rangle \geq \delta_\ell \geq 0, \\
& && \ell = 1, 2, \dots, L, \\
& && \mathbf{X}_1 \succcurlyeq \mathbf{0}, \mathbf{X}_2 \succcurlyeq \mathbf{0}.
\end{aligned}$$

Note that for a complex matrix \mathbf{X} being Hermitian is a requirement for being positive semidefinite. For a simpler notation we define the convex set

$$\mathcal{C} = \{(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^L \times \mathbb{R}^L : u_\ell v_\ell \geq \delta_\ell > 0, u_\ell \geq 0\}. \quad \text{(S.12)}$$

In order to derive the ADMM scheme, after introducing new variables, program (S.11) can be written as

$$\begin{aligned}
& \underset{\{\mathbf{X}_i, \mathbf{Z}_i, \mathbf{u}_i\}_{i=1,2}}{\text{minimize}} && \mathbb{I}_{\mathcal{C}}(\mathbf{u}_1, \mathbf{u}_2) + \sum_{j=1}^2 \text{Tr}(\mathbf{X}_j) + \mathbb{I}_+(\mathbf{Z}_j) && \text{(S.13)} \\
& \text{subject to} && u_{j,\ell} = \langle \mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^*, \mathbf{X}_j \rangle, \ell = 1, 2, \dots, L, j = 1, 2, \\
& && \mathbf{X}_j = \mathbf{Z}_j, j = 1, 2,
\end{aligned}$$

where the constraints are reflected in the indicator functions

$$\mathbb{I}_{\mathcal{C}}(\mathbf{u}, \mathbf{v}) = \begin{cases} 0 & (\mathbf{u}, \mathbf{v}) \in \mathcal{C} \\ +\infty & (\mathbf{u}, \mathbf{v}) \notin \mathcal{C} \end{cases}, \quad \mathbb{I}_+(\mathbf{Z}) = \begin{cases} 0 & \mathbf{Z} \succeq \mathbf{0} \\ +\infty & \mathbf{Z} \not\succeq \mathbf{0} \end{cases}.$$

Defining the dual matrices $\mathbf{P}_1, \mathbf{P}_2$ and the dual vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{R}^L$, the augmented Lagrangian for (S.13) takes the form

$$\begin{aligned}
L(\{\mathbf{X}_i, \mathbf{Z}_i, \mathbf{P}_i, \mathbf{u}_i, \boldsymbol{\alpha}_i\}_{i=1,2}) &= \mathbb{I}_{\mathcal{C}}(\mathbf{u}_1, \mathbf{u}_2) + \sum_{j=1}^2 \text{Tr}(\mathbf{X}_j) + \mathbb{I}_+(\mathbf{Z}_j) \\
&+ \frac{\rho_1}{2} \sum_{j=1}^2 \sum_{\ell=1}^L (u_{j,\ell} - \langle \mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^*, \mathbf{X}_j \rangle + \alpha_{j,\ell})^2 \\
&+ \frac{\rho_2}{2} \sum_{j=1}^2 \|\mathbf{X}_j - \mathbf{Z}_j + \mathbf{P}_j\|_F^2. && \text{(S.14)}
\end{aligned}$$

In an ADMM scheme the update for each variable at the k -th iteration is performed by minimizing L with respect to that variable, while fixing the other ones. More specifically, using the superscript (k) to denote the iteration, for $j = 1, 2$ we have the primal updates

$$\mathbf{X}_j^{(k+1)} = \arg \min_{\mathbf{X}_j} \text{Tr}(\mathbf{X}_j) + \frac{\rho_1}{2} \sum_{\ell=1}^L \left(\langle \mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^*, \mathbf{X}_j \rangle - u_{j,\ell}^{(k)} - \alpha_{j,\ell}^{(k)} \right)^2 + \frac{\rho_2}{2} \left\| \mathbf{X}_j - \mathbf{Z}_j^{(k)} + \mathbf{P}_j^{(k)} \right\|_F^2,$$

$$\mathbf{Z}_j^{(k+1)} = \arg \min_{\mathbf{Z}_j} \frac{1}{2} \left\| \mathbf{Z}_j - \mathbf{X}_j^{(k+1)} - \mathbf{P}_j^{(k)} \right\|_F^2 + \mathbb{I}_+(\mathbf{Z}_j),$$

$$\left(\mathbf{u}_1^{(k+1)}, \mathbf{u}_2^{(k+1)} \right) = \arg \min_{\mathbf{u}_1, \mathbf{u}_2} \frac{1}{2} \sum_{j=1}^2 \sum_{\ell=1}^L \left(u_{j,\ell} - \langle \mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^*, \mathbf{X}_j^{(k+1)} \rangle + \alpha_{j,\ell}^{(k)} \right)^2 + \mathbb{I}_{\mathcal{C}}(\mathbf{u}_1, \mathbf{u}_2),$$

and the dual updates

$$\begin{aligned} \alpha_{j,\ell}^{(k+1)} &= \alpha_{j,\ell}^{(k)} + u_{j,\ell}^{(k+1)} - \langle \mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^*, \mathbf{X}_j^{(k+1)} \rangle \\ \mathbf{P}_j^{(k+1)} &= \mathbf{P}_j^{(k)} + \mathbf{X}_j^{(k+1)} - \mathbf{Z}_j^{(k+1)}. \end{aligned}$$

In the sequel we derive closed-form expressions for all the primal updates. To formulate the \mathbf{X} -update, taking the derivative of the objective with respect to \mathbf{X}_j and setting it to zero yields

$$\mathbf{I} + \rho_1 \sum_{\ell=1}^L \left(\langle \mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^*, \mathbf{X}_j^{(k+1)} \rangle - u_{j,\ell}^{(k)} - \alpha_{j,\ell}^{(k)} \right) \mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^* + \rho_2 \left(\mathbf{X}_j^{(k+1)} - \mathbf{Z}_j^{(k)} + \mathbf{P}_j^{(k)} \right) = \mathbf{0},$$

which after vectorizing $\mathbf{X}_j^{(k+1)}$ yields

$$\text{vec}(\mathbf{X}_j^{(k+1)}) = \mathbf{A}_j^{-1} \text{vec} \left(\rho_1 \sum_{\ell=1}^L \left(u_{j,\ell}^{(k)} + \alpha_{j,\ell}^{(k)} \right) \mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^* + \rho_2 \left(\mathbf{Z}_j^{(k)} - \mathbf{P}_j^{(k)} \right) - \mathbf{I} \right),$$

where

$$\mathbf{A}_j = \rho_1 \sum_{\ell=1}^L \text{vec}(\mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^*) \text{vec}(\mathbf{a}_{j,\ell} \mathbf{a}_{j,\ell}^*)^* + \rho_2 \mathbf{I}.$$

Note that \mathbf{A}_j^{-1} only needs to be calculated once throughout the entire process.

The \mathbf{Z} update is basically the projection of a Hermitian matrix onto the PSD cone. Considering the eigen-decomposition of the Hermitian matrix $\tilde{\mathbf{Z}} \in \mathbb{C}^{n \times n}$:

$$\tilde{\mathbf{Z}} = \mathbf{U} \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{U}^*,$$

all eigenvalues are real, and the solution to

$$\underset{\mathbf{Z}}{\text{minimize}} \quad \frac{1}{2} \left\| \mathbf{Z} - \tilde{\mathbf{Z}} \right\|_F^2 + \mathbb{I}_+(\mathbf{Z})$$

is simply $\mathbf{U} \text{diag}(\max(\lambda_1, 0), \dots, \max(\lambda_n, 0)) \mathbf{U}^*$.

Finally, the \mathbf{u} -update step in the proposed ADMM scheme requires a fast formulation of the projection onto the set \mathcal{C} . It is straightforward to see that program

$$\underset{\mathbf{u}_1, \mathbf{u}_2}{\text{minimize}} \quad \frac{1}{2} \sum_{j=1}^2 \sum_{\ell=1}^L (u_{j,\ell} - \xi_{j,\ell})^2 + \mathbb{I}_{\mathcal{C}}(\mathbf{u}_1, \mathbf{u}_2) \quad (\text{S.15})$$

decouples into L distinct programs of the form

$$\underset{u_1, u_2}{\text{minimize}} \quad \frac{1}{2} \sum_{j=1}^2 (u_j - \xi_j)^2 \quad \text{subject to: } u_1 u_2 \geq \delta > 0, \quad u_1 \geq 0. \quad (\text{S.16})$$

Note that since the case $u_1 u_2 = 0$ leads to a trivial argument, we consider the strict inequality $\delta > 0$. In the sequel we focus on addressing (S.16), as solving (S.16) for each component ℓ would deliver the solution to (S.15). We proceed by forming the Lagrangian for the constrained problem (S.16)

$$l(u_1, u_2, \mu_1, \mu_2) = \frac{1}{2} \left\| \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} - \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \right\|^2 + \mu_1 (\delta - u_1 u_2) - \mu_2 u_1.$$

Along with the primal constraints, the Karush-Kuhn-Tucker optimality conditions are

$$\frac{\partial l}{\partial u_1} = u_1 - \xi_1 - \mu_1 u_2 - \mu_2 = 0, \quad (\text{S.17})$$

$$\frac{\partial l}{\partial u_2} = u_2 - \xi_2 - \mu_1 u_1 = 0, \quad (\text{S.18})$$

$$\begin{aligned} \mu_1 &\geq 0, & \mu_1 (\delta - u_1 u_2) &= 0, \\ \mu_2 &\geq 0, & \mu_2 u_1 &= 0. \end{aligned}$$

We now proceed with the possible cases.

Case 1. $\mu_1 = \mu_2 = 0$:

In this case we have $(u_1, u_2) = (\xi_1, \xi_2)$ and this result would only be acceptable when $u_1 u_2 \geq \delta$ and $u_1 \geq 0$.

Case 2. $\mu_1 = 0, u_1 = 0$:

In this case the first feasibility constraint of (S.16) requires that $\delta \leq 0$, which is not a possibility.

Case 3. $\delta - u_1 u_2 = 0, u_1 = 0$:

Similar to the previous case, this cannot happen when $\delta > 0$.

Case 4. $\mu_2 = 0$, $\delta - u_1 u_2 = 0$:

In this case we have $\delta = u_1 u_2$, combining which with (S.18) yields $\delta = (\xi_2 + \mu_1 u_1) u_1$, or

$$\mu_1 = \frac{\delta - \xi_2 u_1}{u_1^2}. \quad (\text{S.19})$$

Similarly, (S.17) yields

$$u_1 = \xi_1 + \mu_1 (\xi_2 + \mu_1 u_1). \quad (\text{S.20})$$

Since the condition $\delta = u_1 u_2$ requires that $u_1 > 0$, μ_1 can be eliminated between (S.19) and (S.20) to generate the following fourth order polynomial equation in terms of u_1 :

$$u_1^4 - \xi_1 u_1^3 + \delta \xi_2 u_1 - \delta^2 = 0.$$

After solving this 4-th order polynomial equation, we pick the real root u_1 which obeys

$$u_1 \geq 0, \quad \delta - \xi_2 u_1 \geq 0. \quad (\text{S.21})$$

Note that the second inequality in (S.21) warrants nonnegative values for μ_1 thanks to (S.19). After picking the right root, we can explicitly obtain μ_1 using (S.20) and calculate the u_2 using (S.18). The resulting (u_1, u_2) pair presents the solution to (S.16), and finding such pair for every ℓ provides the solution to (S.15).

References

- [Koltchinskii and Mendelson(2015)] Vladimir Koltchinskii and Shahar Mendelson. Bounding the smallest singular value of a random matrix without concentration. *International Mathematics Research Notices*, 2015(23):12991–13008, 2015.
- [Mendelson(2014)] Shahar Mendelson. Learning without concentration. In *Conference on Learning Theory*, pages 25–39, 2014.
- [Lecué et al.(2018)Lecué, Mendelson, et al.] Guillaume Lecué, Shahar Mendelson, et al. Regularization and the small-ball method i: sparse recovery. *The Annals of Statistics*, 46(2): 611–641, 2018.
- [Lecué and Mendelson(2017)] Guillaume Lecué and Shahar Mendelson. Regularization and the small-ball method ii: complexity dependent error rates. *The Journal of Machine Learning Research*, 18(1):5356–5403, 2017.
- [Bahmani and Romberg(2017)] Sohail Bahmani and Justin Romberg. Anchored regression: Solving random convex equations via convex programming. *arXiv preprint arXiv:1702.05327*, 2017.

- [McDiarmid(1989)] Colin McDiarmid. On the method of bounded differences. *Surveys in combinatorics*, 141(1):148–188, 1989.
- [van der Vaart and Wellner(1997)] Aad W van der Vaart and Jon A Wellner. Weak convergence and empirical processes with applications to statistics. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 160(3):596–608, 1997.
- [Ledoux and Talagrand(2013)] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.