Off-the-grid Blind Deconvolution and Demixing

Saeed Razavikia[†], Sajad Daei[†], Mikael Skoglund[†], Gabor Fodor[†,♭], Carlo Fischione[†]
[†]School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden
[♭]Ericsson Research, Sweden
Email: {sraz, sajado, skoglund, gaborf, carlofi}@kth.se

Abstract—We consider the problem of GB2D! (GB2D!) in scenarios where multiple users communicate messages through multiple unknown channels, and a single BS! (BS!) collects their contributions. This scenario arises in various communication fields, including wireless communications, the Internet of Things, over-the-air computation, and integrated sensing and communications. In this setup, each user's message is convolved with a multi-path channel formed by several scaled and delayed copies of Dirac spikes. The BS! receives a linear combination of the convolved signals, and the goal is to recover the unknown amplitudes, continuous-indexed delays, and transmitted waveforms from a compressed vector of measurements at the BS!. However, in the absence of any prior knowledge of the transmitted messages and channels, GB2D! is highly challenging and intractable. To address this issue, we assume that each user's message follows a distinct modulation scheme living in a known low-dimensional subspace. By exploiting these subspace assumptions and the sparsity of the multipath channels for different users, we transform the nonlinear GB2D! problem into a matrix tuple recovery problem from a few linear measurements. To achieve this, we propose a semidefinite programming optimization that exploits the specific low-dimensional structure of the matrix tuple to recover the messages and continuous delays of different communication paths from a single received signal at the BS!. Finally, our numerical experiments show that our proposed method effectively recovers all transmitted messages and the continuous delay parameters of the channels with a sufficient number of samples.

Index Terms—Atomic norm minimization, blind channel estimation, blind data recovery, blind deconvolution, blind demixing.

I. Introduction

In the near future, the Internet of Things (IoT) is expected to connect billions of wireless devices, surpassing the capacity of the current fifth-generation (5G) wireless system both technically and economically. One of the primary challenges that 6G, the future wireless communication system, will face is managing the massive number of IoT devices that generate sporadic traffic. As the 6G market grows, this sporadic traffic will significantly increase, and it is generally agreed among communications engineers that the current 5G channel access procedures cannot handle this volume of traffic.

Traditional channel access methods, which rely on classical information and communication theory, require a large number of pilots or training signals to estimate the channel, leading to significant resource waste that does not scale towards IoT requirements. Thus, minimizing the overhead caused by exchanging certain types of training information, such as channel estimation and data slot assignment, is necessary. This is especially critical for communications over dynamic channels, where channel coherence times are short and channel aging effects cause channel information to change rapidly. In these cases, the assumption of block fading [?] no longer holds. One approach to addressing this issue is to incorporate channel aging effects into the channel estimation process to maximize spectral efficiency (see, e.g., [?]), but this requires knowledge of the channel correlation structure at different times, which might be challenging to obtain in general channel environments. Therefore, for situations where a large number of devices transmit small amounts of data sporadically over highly dynamic channels, and the channel correlation structure is unknown, it is critical to avoid transmitting a signal with a huge overhead before the actual data. This raises the question of whether this is feasible.

To facilitate explanation, we consider a scenario where multiple users transmit messages through multiple frequency-selective channels towards a central BS! (as described in Fig. ??). The BS! receives a combined signal comprising contributions from all users, which is then processed through a series of filters (see Fig. ??). The goal is to simultaneously estimate the transmitted messages and channels from the received measurements at the BS!, which is a challenging bilinear problem.

This scenario appears in a variety of applications, including over-the-air computation [?], [?], super-resolution single-molecule imaging [?], [?], [?], [?], [?], multi-user multipath channel estimation [?], [?], blind calibration in multi-channel sampling systems [?], [?], random access [?] and integrated radar sensing and communications [?], [?], [?].

A. Related work

The problem of recovering messages and channels in the model described above falls into the class of blind deconvolution techniques used to solve inverse problems.

Saeed Razavikia and Carlo Fischione acknowledge the support of WASP, SSF SAICOM, VR, EU FLASH project, and Digital Futures. Sajad Daei and Mikael Skoglund were supported in part by Digital Futures. Digital Futures Project PERCy supported the work of Gabor Fodor.

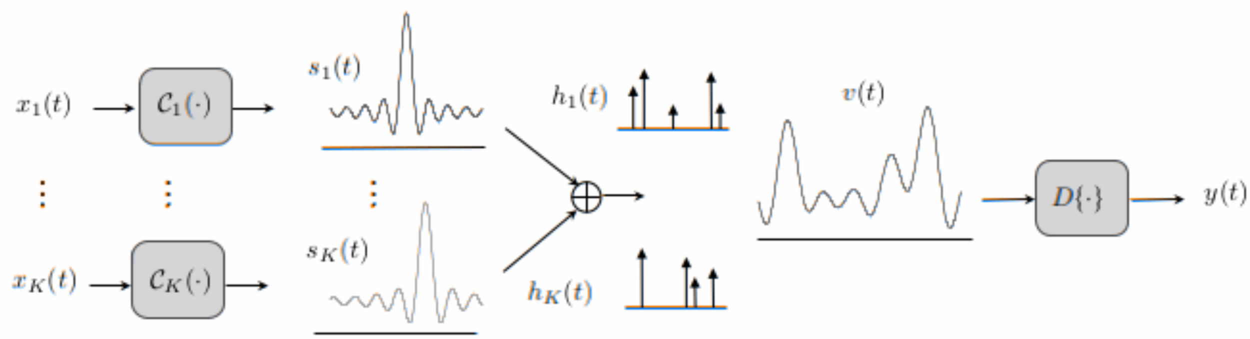


Fig. 1: An illustration of the mathematical model of GB2D! problem. Every user transmits waveform $s_k(t)$ over channel $h_k(t)$ involving P_k multi-path. Afterward, the sum of all these convolved signals is received by $v(t)$.

[illegible]

A tractable convex optimization problem is proposed to recover the continuous channel parameters by promoting the specific feature of the matrix tuple, followed by a least-squares problem to estimate the transmitted messages. In addition, this work provides conditions under which the solution to GB2D! is unique and optimal, and simulation results demonstrate its effectiveness in recovering the channel parameters and transmitted messages with a sufficient number of samples.

Specifically, our contributions are summarized as follows:

- **Blind message recovery and channel estimation with any linear encoder and sensing filter**: **GB2D!** recovers the messages and channels without spending training resources and provides a general communication framework where each user employs a distinct codebook, and the **BS!** employs a linear filter modeling matched filter or sensing block of a communication system.
- **Independent of the number of users**: In a case wherein [illegible]
- **Tractable complexity**: [illegible]
- **Optimality condition**: We provide a theoretical guaran- [illegible]
- **Message and channel distribution are arbitrary**: In [illegible]

[illegible]

Section ??, we provide the GB2D! method, which includes the convex optimization and dual problem to localize the spikes. Section ?? verifies the performance of GB2D! through simulations. Lastly, the paper is concluded in Section ??.

The paper shows vectors and matrices in boldface lower-and upper-case letters, respectively. We use $\boldsymbol{X}^\dagger$ to show the pseudo inverse of matrix $\boldsymbol{X}$. For vector $\boldsymbol{x} \in \mathbb{C}^{N}$ and matrix $\boldsymbol{X} \in \mathbb{C}^{N\times N}$, the norms ℓ_2 is defined as $\|\boldsymbol{x}\|_2 := \sqrt{\sum_{i=1}^{N} |x(i)|^2}$. We further use $\odot$ to show Hadamard products (element-wise products). The Toeplitz lifting operator $\mathscr{T}: \mathbb{C}^N \to \mathbb{C}^{N\times N}$ for given vector $\boldsymbol{x}$ is defined as follows

$$\mathscr{T}(\boldsymbol{x}) := \begin{bmatrix} x_1 & \bar{x}_2 & \cdots & \bar{x}_N \\ x_2 & x_1 & \cdots & \bar{x}_{N-1} \\ \vdots & & & \vdots \\ x_N & x_{N-1} & \cdots & x_1 \end{bmatrix} \quad (1)$$

where the (i,j)-th element is given by $[\mathscr{T}(\boldsymbol{x})]_{i,j} = x_{i-j+1}$ for $i \geq j$ and $[\mathscr{T}(\boldsymbol{x})]_{i,j} = \bar{x}_{j-i+1}$ for $i < j$. Also $\boldsymbol{e}_n$ stands for the n-th column of identity matrix $\boldsymbol{I}_N$. We use $\langle \cdot, \cdot \rangle$ to show the inner product operator where for two arbitrary matrices $\boldsymbol{A}, \boldsymbol{B}$, i.e., $\langle \boldsymbol{A}, \boldsymbol{B} \rangle$ represents $\mathrm{Tr}(\boldsymbol{B}^{\mathsf{H}}\boldsymbol{A})$ and for two continuous functions $f(t)$ and $g(t)$ it means $\int_{-\infty}^{\infty} f(t)g(t)dt$ by $\langle f(t), g(t) \rangle$ based on context. The notation $\boldsymbol{A} \succeq \boldsymbol{0}$ means $\boldsymbol{A}$ is a positive semidefinite matrix.

II. System Model and Problem Formulation

For the purpose of exposition, we consider K single-antenna users transmitting their messages toward a BS! over the MAC! (MAC!). Let $x_k(t)$ be the message of user k, it is mapped into modulated band-limited signal $s_k(t)$ using linear encoder $\mathcal{C}_k(\cdot)$, i.e., $s_k(t) := \mathcal{C}_k(x_k(t))$ for transmitting over the MAC!. Then, all users transmit simultaneously over the same frequency or codes, and the BS! records the signal $v(t)$, which is the sum of all these convolved signals given by

$$v(t) = \sum_{k=1}^{K} h_k(t) \circledast s_k(t), \quad (2)$$

where $h_k(t) = \sum_{\ell=1}^{P_k} g_\ell^k \delta(t - \bar{\tau}_\ell^k)$, is the impulse response of the frequency selective channel corresponding to user k and $\circledast$ is the convolution operator. Furthermore, P_k, $\bar{\tau}_\ell^k$ and g_ℓ^k are the number of multipath delays, the delay, and the complex amplitude of the communication path k corresponding to user k, respectively. The channel delays $\bar{\tau}_\ell^k$'s can take any arbitrary continuous values in $[0,T)$ in which T denotes the duration of the observation time. Afterward, the convoluted signal $v(t)$ goes through linear system $\mathcal{D}(\cdot)$ (e.g. matched-filter or low pass filter) whose output becomes signal $y(t)$, i.e.,

$$y(t) = \mathcal{D}\{v(t)\}, \quad t \in [0,T]. \quad (3)$$

Then, after sampling, the measurements are given by $y_m := y(\frac{m}{T})$ for $m \in [M]$. The goal is to estimate the set of channel delays and amplitudes of $\{h_k(t)\}_{k=1}^{K}$, as well as the unknown transmitted messages $\{x_k(t)\}_{k=1}^{K}$ from the available measurements $y_1, \ldots, y_M$ at the BS! (see Fig ??). We refer to the solution to this problem as GB2D!.

A. Problem formulation

Assuming the measured signal $y(t)$ is square-integrable in Lebesgue's sense (or Band-limited), we can expand it as $y(t) = \sum_{m=1}^{M} y_m \varphi_m(t)$ where φ_m's are compact support basis functions for $t \in [0,T]$ that are orthonormal, i.e.,

$$\langle \varphi_i(t), \varphi_j(t) \rangle := \int_{-\infty}^{\infty} \varphi_i(t)\varphi_j(t)dt = \delta_{i-j}, \quad (4)$$

where δ_{i-j} is the discrete delta Dirac function. Then, m-th sample of the signal in (??) can be written as

$$\begin{aligned} y_m &= \langle \mathcal{D}\{v(t)\}, \varphi_m \rangle \\ &= \langle v(t), \underbrace{\mathcal{D}^*\{\varphi_m\}}_{d_m(t)} \rangle \\ &= \langle \mathscr{F}^*\mathscr{F}\{v(t)\}, d_m(t) \rangle \\ &= \langle \mathscr{F}\{v(t)\}, \mathscr{F}\{d_m(t)\} \rangle \\ &= \langle V(f), D_m(f) \rangle, \end{aligned} \quad (5)$$

where $\mathscr{F}(\cdot)$ denotes linear Fourier operator, and $V(f)$ and $D_m(f)$ are the Fourier transform of $v(t)$ and $d_m(t)$, respectively. Let $\{s_k(t)\}_{k=1}^{K}$ be the transmitted waveforms whose spectrum lie in the interval $[-B, B]$. Taking the Fourier transform of (??); we have the following.

$$V(f) = \sum_{k=1}^{K} H_k(f)S_k(f), \ \forall f \in [-B, B], \quad (6)$$

where $V(f)$, $H_k(f)$, and $S_k(f)$ are the Fourier transform of $v(t)$, $h_k(t)$, and $s_k(t)$, respectively. Substituting (??) into (??), we obtain

$$y_m = \sum_{k=1}^{K} \langle H_k(f)S_k(f), D_m(f) \rangle.$$

By uniformly sampling (??) at N points $f_n = Bn/\lceil (N-1)/2 \rceil$, $n = -\lceil (N-1)/2 \rceil, \ldots, \lceil (N-1)/2 \rceil$ or $n = 0, \ldots, N-1$ provided that $BT \leq \lceil (N-1)/2 \rceil$, we reach

$$y_m = \langle \boldsymbol{d}_m, \sum_{k=1}^{K} \boldsymbol{h}_k \odot \boldsymbol{s}_k \rangle = \langle \boldsymbol{d}_m, \boldsymbol{v} \rangle, \quad (7)$$

where vectors $\boldsymbol{h}_k, \boldsymbol{s}_k, \boldsymbol{d}_m$ and $\boldsymbol{v}$ are defined as

$$\begin{aligned} \boldsymbol{h}_k &= [H(f_1), \ldots, H(f_N)]^{\mathsf{T}}, \\ \boldsymbol{s}_k &= [S(f_1), \ldots, S(f_N)]^{\mathsf{T}}, \\ \boldsymbol{d}_m &= [D_m(f_1), \ldots, D_m(f_N)]^{\mathsf{T}}, \\ \boldsymbol{v} &= [V(f_1), \ldots, V(f_N)]^{\mathsf{T}}. \end{aligned}$$

Note that to set N as small as possible without loss of generality, we choose $N = 2BT + 1$. The relation in (??) can be represented in a matrix form as

$$\boldsymbol{y} = \boldsymbol{D} \sum_{k=1}^{K} \boldsymbol{h}_k \odot \boldsymbol{s}_k, \quad (8)$$

where $\boldsymbol{y} = [y_1, \ldots, y_M]^{\mathsf{T}}$ and $\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M]^{\mathsf{T}} \in \mathbb{C}^{M\times N}$. Now, recall that $h_k(t) = \sum_{\ell=1}^{P_k} g_\ell^k \delta(t - \tau_\ell^k)$ and $s_k(t) = \mathcal{C}_k(x_k(t))$ where $\tau_\ell^k := \bar{\tau}_\ell^k / T$, then we can write

$$\boldsymbol{y} = \boldsymbol{D} \sum_{k=1}^{K} \sum_{\ell=1}^{P_k} g_\ell^k \boldsymbol{a}(\tau_\ell^k) \odot \boldsymbol{C}_k \boldsymbol{x}_k, \quad (9)$$

where $\boldsymbol{a}(\tau) := [1, e^{-j2\pi\tau}, \ldots, e^{-j2\pi(N-1)\tau}]^{\mathsf{T}}$ and

$$\boldsymbol{s}_k = \boldsymbol{C}_k \boldsymbol{x}_k. \quad (10)$$

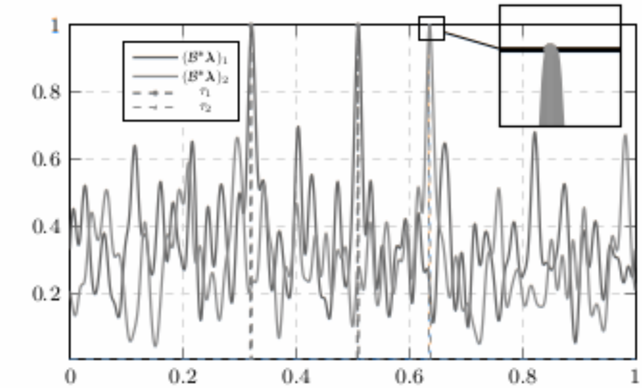


Fig. 2: Delay estimation via dual polynomials with $N = 64$ for $K = 2$ and $P_1 = 2$, $P_2 = M_1$ with $M_1 = M_2 = 5$.

Note that N is set as small as possible without loss of generality. We choose $N = 2B T + 1$. The relation in (??) can be represented in a matrix form as

$$\boldsymbol{y} = \boldsymbol{D}\sum_{k=1}^{K} \boldsymbol{h}_k \odot \boldsymbol{s}_k, \tag{8}$$

where $\boldsymbol{y} = [y_1, \ldots, y_M]^\top$ and $\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M]^\top \in \mathbb{C}^{M\times N}$. Now, recall that $h_k(t) = \sum_{\ell=1}^{P_k} g_\ell^k \delta(t-\tau_\ell^k)$ and $s_k(t) = \mathcal{C}_k(x_k(t))$ where $\tau_\ell^k := \bar{\tau}_\ell^k/T$, then we can write

$$\boldsymbol{y} = \boldsymbol{D}\sum_{k=1}^{K}\sum_{\ell=1}^{P_k} g_\ell^k \boldsymbol{a}(\tau_\ell^k) \odot \boldsymbol{C}_k \boldsymbol{x}_k, \tag{9}$$

where $\boldsymbol{a}(\tau) := [1, e^{-i2\pi\tau}, \ldots, e^{-i2\pi(N-1)\tau}]^\top$ and

$$\boldsymbol{s}_k = \boldsymbol{C}_k \boldsymbol{x}_k. \tag{10}$$

Also, $\boldsymbol{C}_k := [\boldsymbol{c}_1^k, \ldots, \boldsymbol{c}_N^k]^\top \in \mathbb{C}^{N\times M_k}$ is codebook matrix corresponding to encoder $\mathcal{C}_k$ which is a known basis of the subspace with $N \gg M_k$. Further, $\boldsymbol{x}_k \in \mathbb{C}^{M_k}$ is the Fourier transform of message vector of user k. Without loss of generality, we assume that the energy of the message signal is normalized, i.e., $\|\boldsymbol{x}_k\|_2 = 1$ for $k \in [K]$. Our goal is to recover τ_ℓ^ks, g_ℓ^ks, and $\boldsymbol{x}_k$s from the observation vector $\boldsymbol{y} \in \mathbb{C}^M$. Note that it is unavoidable to have phase ambiguities for recovering $\boldsymbol{x}_k$'s and $\boldsymbol{h}_k$'s because of any $\alpha_k \in \mathbb{C}\backslash\{0\}$, we have

$$\boldsymbol{y} = \boldsymbol{D}\sum_{k=1}^{K} \alpha_k \boldsymbol{h}_k \odot \boldsymbol{C}_k \frac{\boldsymbol{x}_k}{\alpha_k}. \tag{11}$$

In the next section, we present the GB2D! method for demixing the measured signals by solving a convex optimization.

III. Proposed Method

In this section, we introduce the main idea GB2D! method and propose a convex optimization to recover all the channel parameters τ_ℓ^ks, $\ell \in [P_k]$ and transmitted waveforms $s_k(t)$s with general known codebook matrices $\boldsymbol{C}_k$'s. Invoking the subspace assumption (??), n-th Fourier samples of signal $v(t)$ can be written as

$$V(f_n) = \sum_{k=1}^{K}\sum_{\ell=1}^{P_k} g_\ell^k \boldsymbol{e}_n^\top \boldsymbol{a}(\tau_\ell^k)\boldsymbol{x}_k^\top \boldsymbol{c}_n^k, \tag{12}$$

where $\boldsymbol{e}_n$ stands for the n-th column of $\boldsymbol{I}_N$. Let $\boldsymbol{X}_k = \sum_{\ell=1}^{P_k} g_\ell^k \boldsymbol{x}_k \boldsymbol{a}(\tau_\ell^k)^\top \in \mathbb{C}^{M_k\times N}$. Using the lifting trick [?], the measurements $V(f_n), n \in [N]$ in (??) can be written as

$$V(f_n) = \sum_{k=1}^{K}\sum_{\ell=1}^{P_k} g_\ell^k \operatorname{Tr}\left(\boldsymbol{c}_n^k \boldsymbol{e}_n^\top \boldsymbol{a}(\tau_\ell^k)\boldsymbol{x}_k^\top\right) = \sum_{k=1}^{K}\langle \boldsymbol{X}_k, \boldsymbol{c}_n^k \boldsymbol{e}_n^\top\rangle.$$

Writing in matrix form, we have $\boldsymbol{v} = \mathcal{C}(\boldsymbol{\mathcal{X}})$, where $\boldsymbol{\mathcal{X}} := (\boldsymbol{X}_k)_{k=1}^K \in \bigoplus_{k=1}^{k} \mathbb{C}^{M_k\times N}$ is the matrix tuple of interest and $\mathcal{C}$ is the linear measurement mapping defined as

$$\bigoplus_{k=1}^{K} \mathbb{C}^{M_k\times N} \to \mathbb{C}^N, \quad (\boldsymbol{Z}_k)_{k=1}^K \to \left(\sum_{k=1}^{K}\langle \boldsymbol{Z}_k, \boldsymbol{c}_n^k \boldsymbol{e}_n^\top\rangle\right)_{n=1}^{N}.$$

Then, by defining $\mathcal{B} := \boldsymbol{D}\mathcal{C}$, the measurements $\boldsymbol{y}$ reads to

$$\boldsymbol{y} = \mathcal{B}(\boldsymbol{\mathcal{X}}). \tag{13}$$

In model (??), the number of delays $\{P_k\}_{k=1}^K$ (e.g., in a multipath channel in multi-user wireless systems) is small. Thus, we define the atomic norm [?]

$$\|\boldsymbol{Z}\|_{\mathcal{A}_k} := \inf\{t > 0 : \boldsymbol{Z} \in t\operatorname{conv}(\mathcal{A}_k)\}$$

$$= \inf_{\substack{c_\ell, \tau_\ell \\ \|\boldsymbol{x}_k\|_2 = 1}} \left\{\sum_\ell |c_\ell| : \boldsymbol{Z} = \sum_\ell c_\ell \boldsymbol{x}_k \boldsymbol{a}(\tau_\ell)^\top \in \mathbb{C}^{M_k\times N}\right\} \tag{14}$$

associated with the atoms

$$\mathcal{A}_k = \{\boldsymbol{x}\boldsymbol{a}(\tau)^\top : \tau \in [0,1), \|\boldsymbol{x}\|_2 = 1, \boldsymbol{x} \in \mathbb{C}^{M_k}\}, k \in [K].$$

The atomic norm $\|\boldsymbol{X}_k\|_{\mathcal{A}_k}$ can be regarded as the best convex alternative for the smallest number of atoms $\mathcal{A}_k$ needed to represent a signal $\boldsymbol{X}_k$. Hence, we are interested in recovering the matrix tuple $\boldsymbol{\mathcal{X}} := (\boldsymbol{X}_k)_{k=1}^K$ by motivating its atomic sparsity by solving the following optimization problem

$$\min_{\boldsymbol{\mathcal{Z}} = (\boldsymbol{Z}_k)_{k=1}^K} \sum_{k=1}^{K} \|\boldsymbol{Z}_k\|_{\mathcal{A}_k} \quad \boldsymbol{y}_{M\times 1} = \mathcal{B}(\boldsymbol{\mathcal{Z}}). \tag{15}$$

Finding the optimal parameters in (??) is not an easy task because it involves an infinite-dimensional variable optimization due to the continuity of the set. Alternatively, we can solve the dual problem explained in the next section.

In what follows, we state the conditions for the uniqueness of the solution to optimization in (??) and the exact recovery of channel parameters and message data. Before expressing Proposition ??, we define the separation between the delays of channel k as

$$\Delta_k := \min_{\ell \neq q} |\tau_\ell^k - \tau_q^k|, \tag{16}$$

and the minimum separation between all users by $\Delta = \min_k \Delta_k$. The absolute value in the latter definition is evaluated as the wrap-around distance on the unit circle.

Proposition 1: *Denote the set of multipath's delays of $h_k(t)$ as $\mathcal{P}_k := \{\tau_\ell^k\}_{\ell=1}^{P_k}$. The solution $\widehat{\boldsymbol{\mathcal{X}}} = (\widehat{\boldsymbol{X}}_k)_{k=1}^K$ of (??) is unique if $\Delta \geqslant \frac{1}{N}$ and there exists a vector $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]^\top \in \mathbb{C}^N$ such that the vector-valued dual polynomials*

$$\boldsymbol{q}_k(\tau) = (\mathcal{B}^*\boldsymbol{\lambda})\boldsymbol{a}(\tau) = \sum_{n=1}^{N} \lambda_n e^{-i2\pi n\tau} \boldsymbol{c}_n^k, \tag{17}$$

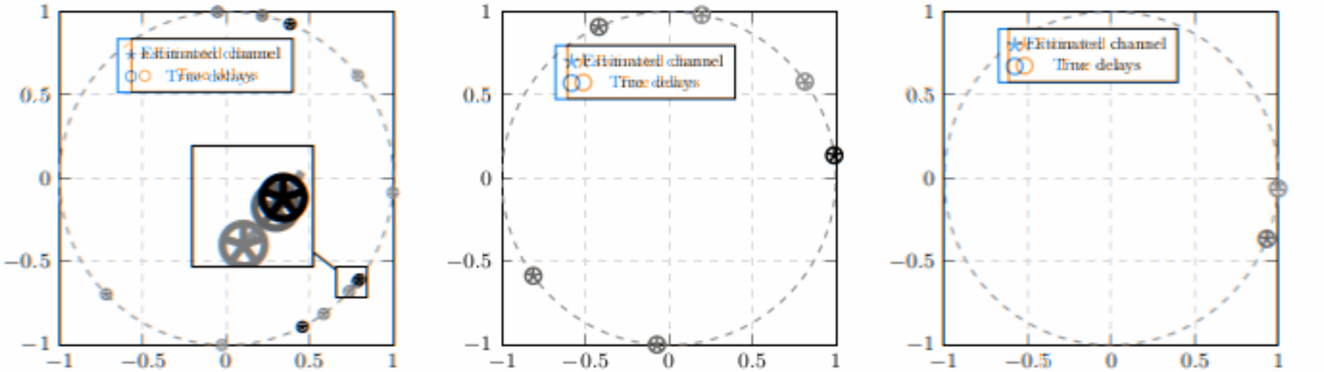
for $k \in [K]$, satisfy the conditions

$$\boldsymbol{q}_k(\tau_\ell^k) = \operatorname{sign}(g_\ell^k)\boldsymbol{x}_k, \quad \tau_\ell^k \in \mathcal{P}_k, k \in [K] \tag{18}$$

$$\|\boldsymbol{q}_k(\tau)\|_2 < 1 \quad \forall \tau \in [0,1)\backslash\mathcal{P}_k, k \in [K]. \tag{19}$$

Proof: See Appendix ??.

Corollary: *The dual polynomial $\boldsymbol{q}_k(\tau)$ only depends on its corresponding codebook, i.e., $\boldsymbol{C}_k$. Therefore, in the case where all users employ the same codebook matrix, all users have a common subspace, i.e., $\boldsymbol{C} \in \mathbb{C}^{N\times M'}$ where M' denotes message size for all the users, all the atomic norms in (??) can be replaced by the same atomic norm and dual polynomial.*



(a) (b) (c)

Fig. 3: Performance of GB2D!. Fig ?? shows the channel estimation for the case where for $K = 4$ users and $P_1 = 3$, $N = 200$ samples and the message size $M_k = 5$. Fig ?? depicts the performance of GB2D! for $K = 2$, $P_1 = 3$, $P_2 = 2$, $N = 128$ samples. Fig ?? shows the output of GB2D! for the case $M_k = 16$ and $N = 64$. Each signal is shown with a different color code.

Proposition 1. Denote the set of multipath's delays of $h_k(t)$ as $\mathcal{P}_k := \{\tau_\ell^k\}_{\ell=1}^{P_k}$. The solution $\mathcal{X} = (\boldsymbol{X}_k)_{k=1}^K$ of (??) is unique if $\Delta \geqslant \frac{1}{N}$ and there exists a vector $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_N]^\mathsf{T} \in \mathbb{C}^N$ such that the vector-valued dual polynomials

$$\boldsymbol{q}_k(\tau) = (\mathcal{B}^*\boldsymbol{\lambda})_k \boldsymbol{a}^*(\tau) = \sum_{n=1}^{N} \lambda_n e^{j2\pi n\tau} \boldsymbol{c}_n^k \in \mathbb{C}^{r_k}, \quad (17)$$

for $k \in [K]$, satisfy the conditions

$$\boldsymbol{q}_k(\tau_\ell) = \mathrm{sgn}(c_\ell)\boldsymbol{x}_k \ \ \forall \tau_\ell \in \mathcal{P}_k, \ k \in [K] \quad (18)$$

$$\|\boldsymbol{q}_k(\tau)\|_2 < 1 \ \ \forall \tau \in [0,1)\backslash\mathcal{P}_k, \ k \in [K]. \quad (19)$$

Proof. See Appendix ??. ■

Corollary. The dual polynomial $\boldsymbol{q}_k(\tau)$ only depends on its corresponding codebook, i.e., $\boldsymbol{C}_k$. Therefore, in the case where all users employ the same codebook matrix, all users have a common subspace, i.e., $\boldsymbol{C}_k = \boldsymbol{C} \in \mathbb{C}^{N\times M'}$ where M' denotes message size for all the users, all the atomic norms in (??) can be replaced with only one atom and its dual polynomial.

A. Channel estimation via Dual Problem

Before proceeding with the dual problem, let us define the dual polynomial of the atomic norm in the sequel. In particular, the dual atomic norm $\|\cdot\|_{\mathcal{A}_k}^{\mathsf{d}}$ at an arbitrary point $\boldsymbol{Z} \in \mathbb{C}^{M_k\times N}$ is defined as

$$\begin{aligned}\|\boldsymbol{Z}\|_{\mathcal{A}_k}^{\mathsf{d}} &:= \sup_{\|\boldsymbol{X}\|_{\mathcal{A}_k}\leq 1} \mathrm{Re}\{\langle \boldsymbol{Z}, \boldsymbol{X}\rangle\} \\ &= \sup_{\tau\in[0,1), \|\boldsymbol{x}\|_2=1} \mathrm{Re}\{\langle \boldsymbol{Z}, \boldsymbol{x}\boldsymbol{a}(\tau)^\mathsf{T}\rangle\} \\ &= \sup_{\tau\in[0,1), \|\boldsymbol{x}\|_2=1} \mathrm{Re}\{\langle \boldsymbol{x}, \boldsymbol{Z}\boldsymbol{a}^*(\tau)\rangle\} \\ &= \sup_{\tau\in[0,1)} \|\boldsymbol{Z}\boldsymbol{a}^*(\tau)\|_2. \end{aligned} \quad (20)$$

Then, by assigning the Lagrangian vector $\boldsymbol{\lambda} \in \mathbb{C}^N$ to the equality constraint of (??), we have

$$\begin{aligned} L(\mathcal{Z},\boldsymbol{\lambda}) &= \inf_{\mathcal{Z}\in\bigoplus_{k=1}^K \mathbb{C}^{M_k\times N}} \Big[\sum_{k=1}^K \|\boldsymbol{Z}_k\|_{\mathcal{A}_k} + \langle \boldsymbol{\lambda}, \boldsymbol{y} - \mathcal{B}(\mathcal{Z})\rangle\Big] \\ &= \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle + \sum_{k=1}^K \inf_{\boldsymbol{Z}_k\in\mathbb{C}^{M_k\times N}} \Big[\|\boldsymbol{Z}_k\|_{\mathcal{A}_k} - \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \boldsymbol{Z}_k\rangle\Big], \end{aligned} \quad (21)$$

where we used that $\langle \mathcal{B}^*\boldsymbol{\lambda}, \mathcal{Z}\rangle = \sum_{k=1}^K \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \boldsymbol{Z}_k\rangle$. By using Hölder's inequality, (??) becomes equivalent to

$$L(\mathcal{Z},\boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle + \sum_{k=1}^K \inf_{\boldsymbol{Z}_k\in\mathbb{C}^{M_k\times N}} \Big[\|\boldsymbol{Z}_k\|_{\mathcal{A}_k}(1 - \|(\mathcal{B}^*\boldsymbol{\lambda})_k\|_{\mathcal{A}_k}^{\mathsf{d}})\Big].$$

Solving the latter optimization problem, we obtain

$$L(\mathcal{Z},\boldsymbol{\lambda}) = \begin{cases} \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle, & \|(\mathcal{B}^*\boldsymbol{\lambda})_k\|_{\mathcal{A}_k}^{\mathsf{d}} \leqslant 1, \ k \in [K] \\ -\infty, & \text{otherwise.} \end{cases} \quad (22)$$

By transforming implicit constraints into explicit ones, the dual problem becomes

$$\max_{\boldsymbol{\lambda}\in\mathbb{C}^N} \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle \quad \text{s.t.} \quad \|(\mathcal{B}^*\boldsymbol{\lambda})_k\|_{\mathcal{A}_k}^{\mathsf{d}} \leqslant 1, \quad k \in [K], \quad (23)$$

where $\mathcal{B}^* : \mathbb{C}^M \to \bigoplus_{k=1}^K \mathbb{C}^{M_k\times N}$ denotes the adjoint operator of $\mathcal{B}$ and $\mathcal{B}^*\boldsymbol{\lambda} := ((\mathcal{B}^*\boldsymbol{\lambda})_k)_{k=1}^K$ is a matrix tuple where the k-th matrix is given by $(\mathcal{B}^*\boldsymbol{\lambda})_k = \sum_{n=1}^N \lambda_n \boldsymbol{c}_n^k \boldsymbol{e}_n^\mathsf{T}$. Maximization in (??) can also be presented in SDP format as

$$\begin{aligned} \boldsymbol{\lambda}^* = \ & \underset{\boldsymbol{\lambda}\in\mathbb{C}^N, \boldsymbol{Q}\in\mathbb{C}^{N\times N}}{\mathrm{argmax}} \ \mathrm{Re}\{\langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle\} \\ & \text{s.t.} \begin{bmatrix} \boldsymbol{Q} & (\mathcal{B}^*\boldsymbol{\lambda})_k^H \\ (\mathcal{B}^*\boldsymbol{\lambda})_k & \boldsymbol{I}_T \end{bmatrix} \succeq \boldsymbol{0}, \ k \in [K] \\ & \langle \mathcal{T}(\boldsymbol{e}_q), \boldsymbol{Q}\rangle = 1_{q=0}, \ q = -N+1, \dots, N-1, \end{aligned} \quad (24)$$

where $\mathcal{T}$ shows the Toeplitz structure. Maximization in (??) is a convex problem; therefore, it can be efficiently solved using the CVX toolbox [?]. Let $\hat{\boldsymbol{\lambda}}$ be the solution to the dual problem in (??), then the spikes can be localized by the peaks of the following term

$$\hat{\tau}_k = \Big\{\tau \in [0,1) \,\Big|\, \|(\mathcal{B}^*\hat{\boldsymbol{\lambda}})_k \boldsymbol{a}(\tau)\|_2 = 1\Big\}.$$

For instance, an example of this channel estimation is depicted in Fig ?? for a case with $K = 2$.

To recover the message vector and the channel amplitudes corresponding to user k, we form $\hat{\boldsymbol{Z}}_k = \sum_\ell \hat{g}_\ell^k \hat{\boldsymbol{x}}_k \boldsymbol{a}(\hat{\tau}_\ell^k)^\mathsf{T}$. Let $\hat{\boldsymbol{g}}^k := [\hat{g}_1^k, \dots, \hat{g}_{P_k}^k]^\mathsf{T}$. Then, the rank one matrix $\hat{\boldsymbol{x}}_k \hat{\boldsymbol{g}}^{k\mathsf{T}}$ can be estimated as $\hat{\boldsymbol{x}}_k \hat{\boldsymbol{g}}^{k\mathsf{T}} = \hat{\boldsymbol{Z}}_k \boldsymbol{A}^{(k)\dagger}$ where $\boldsymbol{A}^{(k)} := [\boldsymbol{a}(\hat{\tau}_1^k), \dots, \boldsymbol{a}(\hat{\tau}_{P_k}^k)]^\mathsf{T}$. By using the assumption $\|\hat{\boldsymbol{x}}_k\|_2 = 1$ and taking singular value decomposition, we can find $\hat{\boldsymbol{x}}_k$ and $\hat{\boldsymbol{g}}^k$ for $k = 1, \dots, K$.

IV. Simulation Results

This section evaluates the estimation performance of GB2D! for different channel delays and messages lengths. The numerical experiments are implemented using MATLAB CVX Toolbox [?]. The delays' locations are generated uniformly at random with the minimum separation $\Delta \geqslant \frac{1}{N}$ to be smaller than what we theoretically expected. The basis of low dimensional tall matrix $\boldsymbol{C}_k$ is generated uniformly at random for $k \in [K]$ from normal distribution $\mathcal{N}(0,1)$. The messages $\boldsymbol{x}_k, k = 1, \dots, K$ are generated i.i.d and uniformly at random from the unit sphere. Note that if

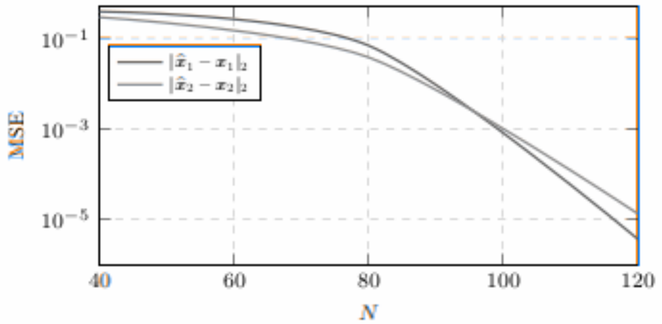


Fig. 4: Message recovery performance versus the number of samples N. Here, both transmitted signals have a message of size 4, i.e., $M_1 = M_2 = 4$. Moreover, we consider the number of channel multipath components as $P_1 = 5$ and $P_2 = 5$.

solved using the CVX toolbox [?]. Let $\hat{\lambda}$ be the solution to the dual problem in (??), then the spikes can be localized by the peaks of the following term $\hat{\mathcal{T}}_k = \{\tau \in [0,1) \mid \|(\mathcal{B}^*\hat{\lambda})_k \boldsymbol{a}(\tau)\|_2 = 1\}$. For instance, an example of this channel estimation is depicted in Fig. ?? for a case with $K = 2$.

To recover the message vector and the channel amplitudes corresponding to user k, we form $\hat{\boldsymbol{Z}}_k = \sum_\ell \hat{g}_\ell^k \hat{\boldsymbol{x}}_k \boldsymbol{a}(\hat{\tau}_\ell^k)^\top$. Let $\hat{\boldsymbol{g}}^k := [\hat{g}_1^k, \ldots, \hat{g}_{P_k}^k]^\top$. Then, the rank one matrix $\hat{\boldsymbol{x}}_k$ can be estimated as $\hat{\boldsymbol{x}}_k \hat{\boldsymbol{g}}^{k\top} = \hat{\boldsymbol{Z}}_k \boldsymbol{A}^{(k)\dagger}$ where $\boldsymbol{A}^{(k)} := [\boldsymbol{a}(\hat{\tau}_1^k), \ldots, \boldsymbol{a}(\hat{\tau}_{P_k}^k)]^\top$. By using the assumption $\|\hat{\boldsymbol{x}}_k\|_2 = 1$ and taking singular value decomposition, we can find $\hat{\boldsymbol{x}}_k$ and $\hat{\boldsymbol{g}}^k$ for $k = 1, \ldots, K$.

IV. Simulation Results

This section evaluates the optimization performance in (??) for different channel delays and message lengths. Then numerical experiments are implemented using MATLAB CVX Toolbox [?]. The delays' locations are generated uniformly at random with the minimum separation $\Delta \geqslant \frac{1}{N}$ to be smaller than what we theoretically expected. The basis of low dimensional tall matrix $\boldsymbol{C}_k \in \mathbb{C}^{N \times M_k}$ is generated uniformly at random for $k \in [K]$ from normal distribution $\mathcal{N}(0,1)$. The messages $\boldsymbol{x}_k, k = 1, \ldots, K$ are generated i.i.d and uniformly at random from the unit sphere. Note that if there is some sort of coordination on the values of transmitted messages (e.g., positiveness) between users and BS!, GB2D! can unambiguously recover the transmitted messages.

For the first case, we set $K = 4$ with $P_1 = \cdots = P_4 = 3$ from $N = 200$ samples, and the filter size $M_k = 5$ for $k \in [4]$. Also, the sensing matrix $\boldsymbol{D}$ is set to be identity, i.e., $\boldsymbol{D} = \boldsymbol{I}_N$. Then, the results are depicted in Fig ??. Fig ?? shows that the GB2D! can distinguish two closely spaced delays. In Fig ??, we repeat this experiment for $K = 3$ with $M_1 = 3, M_2 = 2, M_3 = 1$ from $N = 128$ samples.

For the last case, we check the stability of the algorithm regarding a larger message size $M_k = 16$ with $N = 128$ samples and the effect of the sensing matrix with $M = 64$ (sub-sampling uniform) in Fig ??. We observe that all the cases are successfully recovered by GB2D! for a sufficient number of samples N.

Finally, we evaluate the performance of the GB2D! method for message recovery in Fig ??. Fig ?? depicts the mean square error (MSE) of the estimated messages by the GB2D! for different numbers of samples N. We generate two messages of size 4, i.e., $M_1 = M_2 = 4$ with positive elements. The channel amplitudes are generated as complex Gaussian distribution, and the number of multipath components is set to $P_1 = 5$ and $P_2 = 5$. As it turns out from Fig. ??, GB2D! provides excellent performance in message recovery, and the messages are unambiguously estimated[1]. Moreover, increasing the number of samples at the BS! leads to higher message recovery performance.

The results show that for different numbers of multipath components and various sizes of messages, GB2D! can simultaneously recover messages and estimate multipath channels.

V. Conclusion

The focus of our paper was to explore the possibility of simultaneous data recovery and channel estimation when multiple users send their messages through multiple channels and a BS! receives a linear combination of multiple convolved channels that are made up of a few scaled and delayed continuous Dirac spikes. Specifically, the measurements we worked with were a linear combination of the collective sum of these convolved signals with unknown amplitudes. The main goal was to find these unknown channel delay parameters from only one vector of observations. Since this problem is inherently highly challenging to solve, we overcome this issue by restricting the domain of the transmitted signals to some known low-dimensional subspaces while we have a separation condition on the Dirac spikes. Afterward, we proposed a semidefinite programming optimization to recover the channel delays and the messages of different users simultaneously from one observed vector. For future works, we will provide the performance guarantee of GB2D! by obtaining the required sample complexity that one needs for perfect recovery of the continuous parameters and the transmitted waveforms. Moreover, we plan to investigate further the impact of noise on the GB2D!'s performance and explore possible techniques to enhance its robustness.

Appendix

Any $\boldsymbol{\lambda} \in \mathbb{C}^N$ satisfying (??) and (??) is a feasible point in the dual problem (??). Recall that $\boldsymbol{\mathcal{X}} = (\boldsymbol{X}_k)_{k=1}^K$ is the matrix tuple of interests where $\boldsymbol{X}_k = \sum_{\ell=1}^{P_k} g_\ell^k \boldsymbol{x}_k \boldsymbol{a}(\tau_\ell^k)^\top$. It holds that

$$\sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k} \geqslant \sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k} \|(\mathcal{B}^*\boldsymbol{\lambda})_k\|_{\mathcal{A}_k^*}$$
$$\geqslant \sum_{k=1}^K \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \boldsymbol{X}_k \rangle$$
$$= \sum_{k=1}^K \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \sum_{\ell} g_\ell^k \boldsymbol{x}_k \boldsymbol{a}(\tau_\ell^k)^\top \rangle$$
$$= \sum_{k=1}^K \sum_{\ell=1}^{P_k} \mathrm{Re}\left\{ g_\ell^{k*} \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \boldsymbol{x}_k \boldsymbol{a}(\tau_\ell^k)^\top \rangle \right\}$$
$$= \sum_{k=1}^K \sum_{\ell=1}^{P_k} \mathrm{Re}\left\{ g_\ell^{k*} \langle \boldsymbol{Q}_k(\tau_\ell^k), \boldsymbol{x}_k \rangle \right\} \tag{25}$$

where the second inequality is due to Hölder's inequality, and the last equalities are due to the definition of $\boldsymbol{Q}_k(\tau_\ell^k)$ in (??). We proceed (??) by using conditions (??) and (??):

$$\sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k} \geqslant \sum_{k=1}^K \sum_{\ell=1}^{P_k} \mathrm{Re}\left\{ g_\ell^{k*} \frac{1}{\|\boldsymbol{x}_k\|_2^2} \mathrm{sgn}(g_\ell^k) \langle \boldsymbol{x}_k, \boldsymbol{x}_k \rangle \right\}$$
$$= \sum_{k=1}^K \sum_{\ell=1}^{P_k} |g_\ell^k|$$
$$\geqslant \sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k}, \tag{26}$$

where we used the definition of atomic norm (??) in the last step. From (??) and (??), we find that $\langle \boldsymbol{\lambda}, \mathcal{B}\boldsymbol{\mathcal{X}} \rangle = \sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k}$. Since the pair $(\boldsymbol{\mathcal{X}}, \boldsymbol{\lambda})$ is primal-dual feasible, we reach the conclusion that $\boldsymbol{\mathcal{X}}$ is an optimal solution of (??) and $\boldsymbol{\lambda}$ is an optimal solution of (??) by strong duality. For

[1] Note that the positiveness of the message elements and the ℓ_2 normalized assumption of the message vector, i.e., $\|\boldsymbol{x}_k\|_2 = 1, k = 1, \ldots, K$ are one of the ways to remove the multiplicative ambiguity caused by multiplying channels and the messages.

matrix tuple of interest where $\mathcal{X} = (\boldsymbol{X}_k)_{k=1}^K$ ... It holds that

$$\sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k} \geqslant \sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k} \|(\mathcal{B}^*\boldsymbol{\lambda})_k\|_{\mathcal{A}_k}^d$$
$$\geqslant \sum_{k=1}^K \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \boldsymbol{X}_k \rangle$$
$$= \sum_{k=1}^K \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \sum_{\ell=1}^{P_k} g_\ell^k \boldsymbol{x}_k \boldsymbol{a}(\tau_\ell^k)^\mathsf{T} \rangle$$
$$= \sum_{k=1}^K \sum_{\ell=1}^{P_k} \mathrm{Re}\Big\{ g_\ell^{k*} \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \boldsymbol{x}_k \boldsymbol{a}(\tau_\ell^k)^\mathsf{T} \rangle \Big\}$$
$$= \sum_{k=1}^K \sum_{\ell=1}^{P_k} \mathrm{Re}\Big\{ g_\ell^{k*} \langle \boldsymbol{Q}_k(\tau_\ell^k), \boldsymbol{x}_k \rangle \Big\}, \tag{25}$$

where the second inequality is due to Hölder's inequality, and the last equalities are due to the definition of $\boldsymbol{Q}_k(\tau_\ell^k)$ in (??). We proceed (??) by using conditions (??) and (??):

$$\sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k} \geqslant \sum_{k=1}^K \sum_{\ell=1}^{P_k} \mathrm{Re}\Big\{ g_\ell^{k*} \langle \frac{1}{\|\boldsymbol{x}_k\|_2^2} \mathrm{sgn}(g_\ell^k) \boldsymbol{x}_k, \boldsymbol{x}_k \rangle \Big\}$$
$$= \sum_{k=1}^K \sum_{\ell=1}^{P_k} |g_\ell^k|$$
$$\geqslant \sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k}, \tag{26}$$

where we used the definition of atomic norm (??) in the last step. From (??) and (??), we find that $\langle \boldsymbol{\lambda}, \mathcal{B}\mathcal{X} \rangle = \sum_{k=1}^K \|\boldsymbol{X}_k\|_{\mathcal{A}_k}$. Since the pair $(\mathcal{X}, \boldsymbol{\lambda})$ is primal-dual feasible, we reach the conclusion that $\mathcal{X}$ is an optimal solution of (??) and $\boldsymbol{\lambda}$ is an optimal solution of (??) by strong duality. For proving uniqueness, suppose $\widehat{\mathcal{X}} := (\widehat{\boldsymbol{X}}_k)_{k=1}^K$ is another optimal solution of (??) where $\widehat{\boldsymbol{X}}_k = \sum_{\widehat{\tau}_\ell^k \in \widehat{\mathcal{P}}_k} \widehat{g}_\ell^k \widehat{\boldsymbol{x}}_k \boldsymbol{a}(\widehat{\tau}_\ell^k)^\mathsf{T}$. If $\widehat{\mathcal{X}}$ and $\mathcal{X}$ have the same set of delays, i.e., $\widehat{\mathcal{P}}_k = \mathcal{P}_k,\ \forall\, k \in [K]$, we then have $\widehat{\mathcal{X}} = \mathcal{X}$ since the set of atoms building $\mathcal{X}$ are linearly independent. If there exists some $\widehat{\tau}_\ell^k \notin \mathcal{P}_k$, then we can expand term $\langle \boldsymbol{\lambda}, \mathcal{B}\widehat{\mathcal{X}} \rangle$ as follows

$$\sum_{k=1}^K \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \widehat{\boldsymbol{X}}_k \rangle = \sum_{k=1}^K \sum_{k} \mathrm{Re}\Big\{ \widehat{g}_\ell^{k*} \langle (\mathcal{B}^*\boldsymbol{\lambda})_k, \widehat{\boldsymbol{x}}_k \boldsymbol{a}(\widehat{\tau}_\ell^k)^\mathsf{T} \rangle \Big\} =$$
$$\sum_{k=1}^K \Big[\sum_{\widehat{\tau}_\ell^k \in \mathcal{P}_k} \mathrm{Re}\Big\{ \widehat{g}_\ell^{k*} \langle \boldsymbol{Q}_k(\widehat{\tau}_\ell^k), \widehat{\boldsymbol{x}}_k \rangle \Big\} + \sum_{\widehat{\tau}_\ell^k \notin \mathcal{P}_k} \mathrm{Re}\Big\{ \widehat{g}_\ell^{k*} \langle \boldsymbol{Q}_k(\widehat{\tau}_\ell^k), \widehat{\boldsymbol{x}}_k \rangle \Big\} \Big] \leqslant$$
$$\sum_{k=1}^K \Big[\sum_{\widehat{\tau}_\ell^k \in \mathcal{P}_k} |\widehat{g}_\ell^k| \|\boldsymbol{Q}_k(\widehat{\tau}_\ell^k)\|_2 \|\widehat{\boldsymbol{x}}_r\|_2 + \sum_{\widehat{\tau}_\ell^k \notin \mathcal{P}_k} |\widehat{g}_\ell^k| \|\boldsymbol{Q}_k(\widehat{\tau}_\ell^k)\|_2 \|\widehat{\boldsymbol{x}}_k\|_2 \Big]$$
$$< \sum_{k=1}^K \Big[\sum_{\widehat{\tau}_\ell^k \in \mathcal{P}_k} |\widehat{g}_\ell^k| + \sum_{\widehat{\tau}_\ell^k \notin \mathcal{P}_k} |\widehat{g}_\ell^k| \Big]$$
$$= \sum_{k=1}^K |\widehat{g}_\ell^k| = \sum_{k=1}^K \|\widehat{\boldsymbol{X}}_k\|_{\mathcal{A}_k}, \tag{27}$$

where we used the conditions (??) and (??). The relation (??) contradicts strong duality; hence $\mathcal{X}$ is the unique optimal solution of (??).

proving uniqueness, suppose $\widehat{\mathcal{X}} := (\widehat{\boldsymbol{X}}_k)_{k=1}^K$ is another optimal solution of (??) where $\widehat{\boldsymbol{X}}_k = \sum_{\widehat{\tau}_\ell^k \in \widehat{\mathcal{P}}_k} \widehat{g}_\ell^k \widehat{\boldsymbol{x}}_k \boldsymbol{a}(\widehat{\tau}_\ell^k)^\mathsf{T}$. If $\widehat{\mathcal{X}}$ and $\mathcal{X}$ have the same set of delays, i.e., $\widehat{\mathcal{P}}_k = \mathcal{P}_k,\ \forall\, k \in [K]$, we then have $\widehat{\mathcal{X}} = \mathcal{X}$ since the set of atoms building $\mathcal{X}$ are linearly independent. If there exists some $\widehat{\tau}_\ell^k \notin \mathcal{P}_k$, then we can expand term $\langle \boldsymbol{\lambda}, \mathcal{B}\widehat{\mathcal{X}} \rangle$ as follows

where we used the conditions (??) and (??). The relation (??) contradicts strong duality; hence $\mathcal{X}$ is the unique optimal solution of (??).