Signal Processing 160 (2019) 1–12

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Grab-n-Pull: A max-min fractional quadratic programming framework with applications in signal and information processing☆

Ahmad Gharanjik[a,*], Mojtaba Soltanalian[b], M. R. Bhavani Shankar[a], Björn Ottersten[a]

[a] Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, L-2721, Luxembourg
[b] Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 United States

ARTICLE INFO

Article history:
Received 18 September 2018
Revised 12 January 2019
Accepted 5 February 2019
Available online 7 February 2019

Keywords:
Beamforming
Fractional programming
Max-min optimization
Non-convex optimization
Robust signal processing
Signal design

ABSTRACT

In this work, we propose and study a novel optimization framework (which we call *Grab-n-Pull*, or in abbreviated form, *GnP*) for signal design problems involving max-min fractional quadratic programming. Such optimization problems occur frequently in various subareas of signal and information processing. Using a penalized version of the original design problem, we derive a simplified quadratic reformulation of the problem in terms of the signal (to be designed). Each iteration of the proposed design framework consists of a combination of power method-like iterations and the Gram-Schmidt process, and as a result, enjoys a low computational cost. Particularly, the numerical examples show that the proposed method outperforms the widely used semidefinite relaxation approach in terms of both the quality of approximate solutions and the computational cost. Moreover, the suggested approach can handle various types of signal constraints such as total-power, per-antenna power, unimodularity, or discrete-phase requirements—an advantage which is not shared by other existing approaches in the literature.

Published by Elsevier B.V.

1. Introduction

Notation: We use bold lowercase letters for vectors and bold uppercase letters for matrices. Please see Table 1 for other notations used throughout this paper.

Maximizing the minimal performance is a widely used proactive approach to achieve *fairness* [2–11], *robustness* [12–18], or *efficiency* [19–23] in networked systems requiring advanced signal processing. Interestingly, many of such applications share a similar structure of the performance metric; namely, a variety of quality metrics for signal design, including e.g. signal-to-noise (plus interference) ratio (SINR) and mean-square error (MSE), can be represented as a ratio of quadratic functions of the signal to be designed—several examples will be presented shortly in Section 1.1. The goal of this paper is therefore to study and propose an efficient approach to signal design dealing with the following NP-hard [7,8] optimization problem:

$$\mathcal{P}_1: \quad \max_{\boldsymbol{w}} \quad \min_{i\in[K]} \left\{ \frac{\boldsymbol{w}^H \boldsymbol{A}_i \boldsymbol{w}}{\boldsymbol{w}^H \boldsymbol{B}_i \boldsymbol{w}} \right\}, \quad \text{s. t.} \quad \boldsymbol{w} \in \Omega \tag{1}$$

where $\boldsymbol{w} \in \mathbb{C}^N$ is the signal to be designed, $\boldsymbol{A}_i \in \mathbb{C}^{N\times N}$ and $\boldsymbol{B}_i \in \mathbb{C}^{N\times N}$ are positive semidefinite (PSD) matrices, and Ω denotes the feasible set of $\boldsymbol{w}$ determined by the associated signal constraints. For instance the feasible set can be comprised of unimodular vector $\boldsymbol{w}$ or vector $\boldsymbol{w}$ with a finite-energy, depending on the application.

1.1. Applications in signal and information processing

We describe below several examples from signal processing applications that require tackling $\mathcal{P}_1$.

• *Precoding for fairness-achieving networks:*

A common interpretation of fairness in the networks entails allocating the available resources in order to maximize the minimal user performance [2–10]. In such scenarios, a judicious design of the *precoding* signals for different users can be viewed as a vital part of the network configuration. We consider the general *multigroup multicast* precoding problem [8] for a downlink channel, with a n_{Tx}-antenna transmitter and K single-antenna users assigned to $G \leq K$ multicast groups. In multigroup multicast scenario, G different streams are directed to K users, each with their own channel.

☆ This work was supported in part by European Research Council (ERC) Advanced Grant AGNOSTIC-742648, U.S. National Science Foundation (NSF) Grants CCF-1704401, ECCS-1809225 and the National Research Fund (FNR) of Luxembourg, under AFR Grant 5779106. Some parts of this work were presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2016 [1].

* Corresponding author.
E-mail address: ahmad.gharanjik@databourg.com (A. Gharanjik).

https://doi.org/10.1016/j.sigpro.2019.02.006
0165-1684/Published by Elsevier B.V.

Table 1
Notations.

$\boldsymbol{x}(k)$	The kth entry of the vector $\boldsymbol{x}$
$\|\boldsymbol{x}\|_n$	The l_n-norm of $\boldsymbol{x}$, defined as $(\sum_k \|\boldsymbol{x}(k)\|^n)^{\frac{1}{n}}$
$\boldsymbol{X}^H$	The complex conjugate of a matrix $\boldsymbol{X}$
$\boldsymbol{X}^T$	The transpose of a matrix $\boldsymbol{X}$
$\boldsymbol{X}^{\frac{1}{2}}$	The Hermitian square-root of $\boldsymbol{X}$
$\mathrm{Tr}(\boldsymbol{X})$	The trace of a matrix $\boldsymbol{X}$
$\|\boldsymbol{X}\|_F$	The Frobenius norm of a matrix $\boldsymbol{X}$
$\mathrm{vec}(\boldsymbol{X})$	The vector obtained by column-wise stacking of $\boldsymbol{X}$
$\arg(\boldsymbol{X})$	The phase angle (in radians) of $\boldsymbol{X}$
$\sigma_{\max}(\boldsymbol{X})$	The maximal eigenvalue of $\boldsymbol{X}$
$\sigma_{\min}(\boldsymbol{X})$	The minimal eigenvalue of $\boldsymbol{X}$
$\Re\{\boldsymbol{X}\}$	The real part of $\boldsymbol{X}$
$\mathbb{E}\{\boldsymbol{X}\}$	The expected value of the matrix random variable $\boldsymbol{X}$
$\mathrm{Diag}(\boldsymbol{x})$	The diagonal matrix formed by the entries of $\boldsymbol{x}$
$\boldsymbol{X}\succeq\boldsymbol{Y}$	The matrix $\boldsymbol{X}-\boldsymbol{Y}$ is positive semidefinite definite
$\otimes$	The Kronecker product of matrices
$\boldsymbol{I}_n$	The identity matrix of dimension n
$\boldsymbol{e}_i$	The ith standard basis vector in $\mathbb{C}^K$
$\mathbb{R}_+$	The set of positive real numbers
$\mathbb{C}$	The set of complex numbers
$[K]$	The set $\{1,2,\ldots,K\}$

We denote the subset of user indices in the k^{th} group by $\mathcal{G}_k$ for any $k\in[G]$. Let $\boldsymbol{h}_i\in\mathbb{C}^{n_{\mathrm{Tx}}}$ denote the channel between the transmit antennas and the ith user. Also let $\boldsymbol{w}_k\in\mathbb{C}^{n_{\mathrm{Tx}}}$ denote the *precoding* vector corresponding to the kth, $k\in[G]$, multicast group of users. To form the data stream to the users, any complex symbol to be transmitted, will be modulated by the precoding vector of the intended group of users. The signal transmitted to the group k from the antenna array takes the form $\sum_{k=1}^{G}\boldsymbol{w}^H s_k(t)$, where $s_k(t)$ is the information stream to the users in kth multicast group. Further $s_k(t)$ is modelled as a random variable with zero-mean and unit variance. The precoding vectors are to be designed in order to enhance the network performance. In particular, the SINR value for any user $i\in\mathcal{G}_k$ (and any $k\in[G]$) is given by Karipidis et al. [8,24]

$$\mathrm{SINR}_i=\frac{\boldsymbol{w}_k^H\boldsymbol{R}_i\boldsymbol{w}_k}{\left(\sum_{j\in[G]\setminus\{k\}}\boldsymbol{w}_j^H\boldsymbol{R}_i\boldsymbol{w}_j\right)+\sigma_i^2},\tag{2}$$

where $\boldsymbol{R}_i=\mathbb{E}\{\boldsymbol{h}_i\boldsymbol{h}_i^H\}$ is the covariance matrix of the ith channel, σ_i^2 denotes the variance of the zero-mean additive white Gaussian noise (AWGN).

Consequently, the problem of maximizing the minimal user SINR performance in the network can be formulated as [8],

$$\max_{\{\boldsymbol{w}_k\}_{k=1}^G}\ \min_{k\in[G]}\left\{\ \min_{i\in\mathcal{G}_k}\left\{\frac{\boldsymbol{w}_k^H\boldsymbol{R}_i\boldsymbol{w}_k}{\left(\sum_{j\in[G]\setminus\{k\}}\boldsymbol{w}_j^H\boldsymbol{R}_i\boldsymbol{w}_j\right)+\sigma_i^2}\right\}\right\}$$
$$\text{s. t.}\quad\sum_{k=1}^{G}\|\boldsymbol{w}_k\|_2^2\le P.\tag{3}$$

Note that by a specific reformulation, the SINR metric in (2) can be rewritten as a fractional quadratic criterion. To see this, define the *stacked* precoding vector $\boldsymbol{w}\in\mathbb{C}^N$ (with $N=n_{\mathrm{Tx}}G$) as

$$\boldsymbol{w}\triangleq\mathrm{vec}([\boldsymbol{w}_1\ \boldsymbol{w}_2\ \ldots\ \boldsymbol{w}_G]),\tag{4}$$

and observe that (2) will increase for any increased scaling of $\boldsymbol{w}$. As a result, any finite-energy constraint on $\boldsymbol{w}$ while maximizing $\{\mathrm{SINR}_i\}$ will be *active*, i.e. it will be satisfied with equality. Accordingly, we let $\|\boldsymbol{w}\|_2^2=P$, implying the constraint is active and $\boldsymbol{w}$ should be designed with maximum possible energy. The we make use of the definitions,

$$\boldsymbol{A}_i\triangleq\mathrm{Diag}(\boldsymbol{e}_i)\otimes\boldsymbol{R}_i,\quad\forall i\in[K],\tag{5}$$

$$\boldsymbol{B}_i\triangleq(\boldsymbol{I}_K-\mathrm{Diag}(\boldsymbol{e}_i))\otimes\boldsymbol{R}_i+\frac{\sigma_i^2}{P}\boldsymbol{I}_N,\quad\forall i\in[K].\tag{6}$$

Now, it is not difficult to verify that

$$\mathrm{SINR}_i=\frac{\boldsymbol{w}^H\boldsymbol{A}_i\boldsymbol{w}}{\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w}},\quad\forall i\in[K],\tag{7}$$

in which $\{\boldsymbol{A}_i\}$ are PSD and $\{\boldsymbol{B}_i\}$ are positive definite (PD). As a result, the precoding design problem for maximizing the minimal user SINR performance can be formulated as $\mathcal{P}_1$. Note that $\mathcal{P}_1$ may also be used to formulate the weighted SINR optimization problems; see [2,3,8,25,26] for details.

The problem of optimizing a particular SINR_i in Eq. (7) takes the form of a general Rayleigh coefficient maximization. However, unlike the Rayleigh coefficient optimization, the problem pursued involved constraints on the feasible set of $\boldsymbol{w}$ and also involves optimization of multiple coefficients.

• *Relay beamforming:*

Relays are typically needed to improve the communication performance between user pairs experiencing poor channel quality. We consider a MIMO AF two-way relay network consisting of M_R antennas, L operators and pairs of user terminals as described in [27,28]. We assume single-antenna user terminals and flat fading channels between the ith user of the jth operator and the relay, which are denoted by $\{\boldsymbol{h}_{i,j}\}$ [29]. The received signal at the relay can be expressed as [27–29],

$$\boldsymbol{r}=\sum_{j=1}^{L}\sum_{i=1}^{2}\boldsymbol{h}_{i,j}x_{i,j}+\boldsymbol{n}_R,\tag{8}$$

where $x_{i,j}$ is the transmitted symbol by the ith user of the jth operator with power $p_{i,j}$ (given by $\mathbb{E}\{|x_{i,j}|^2\}$), and $\boldsymbol{n}_R$ denotes the circularly symmetric white Gaussian noise with covariance matrix $\sigma_R^2\boldsymbol{I}$ at the relay. By employing the AF protocol, the transmit signal of the relay is given by

$$\widetilde{\boldsymbol{r}}=\boldsymbol{W}\boldsymbol{r}\tag{9}$$

with $\boldsymbol{W}\in\mathbb{C}^{M_R\times M_R}$ being the relay amplification matrix, which is to be designed. Assuming channel reciprocity between the relay and users [27], the received signal $y_{i,j}$ of the ith user at the jth operator becomes

$$y_{i,j}=\boldsymbol{h}_{i,j}^T\widetilde{\boldsymbol{r}}+n_{i,j},\tag{10}$$

where $n_{i,j}$ is the associated (white) noise component (with variance $\sigma_{i,j}^2$). The minimal user-rate in the network can be formulated as [27]

$$R_{\min}=\frac{1}{2}\min_{j\in[L],\,i\in[2]}\log_2(1+\gamma_{i,j}).\tag{11}$$

Herein, $\gamma_{i,j}$ denotes the signal-to-interference-plus-noise ratio (SINR) for the ith user of the jth operator and it has the following expression [27]

$$\gamma_{i,j}=\frac{\boldsymbol{w}^H\boldsymbol{\Phi}_{i,j}\boldsymbol{w}}{\boldsymbol{w}^H(\boldsymbol{\Upsilon}_{i,j}+\boldsymbol{\Delta}_{i,j})\boldsymbol{w}+\sigma_{i,j}^2},\tag{12}$$

where $\boldsymbol{w}=\mathrm{vec}(\boldsymbol{W})$ and the matrices $\boldsymbol{\Phi}_{i,j}$, $\boldsymbol{\Upsilon}_{i,j}$, $\boldsymbol{\Delta}_{i,j}$ are defined as

$$\boldsymbol{\Phi}_{i,j}=p_{i,j}\left(\boldsymbol{h}_{3-i,j}^T\otimes\boldsymbol{h}_{i,j}^T\right)^H\left(\boldsymbol{h}_{3-i,j}^T\otimes\boldsymbol{h}_{i,j}^T\right),\tag{13}$$
$$\boldsymbol{\Upsilon}_{i,j}=\sum_{\widetilde{i}}\sum_{\widetilde{j}\neq j}p_{\widetilde{i},\widetilde{j}}\left(\boldsymbol{h}_{\widetilde{i},\widetilde{j}}^T\otimes\boldsymbol{h}_{i,j}^T\right)^H\left(\boldsymbol{h}_{\widetilde{i},\widetilde{j}}^T\otimes\boldsymbol{h}_{i,j}^T\right),$$
$$\boldsymbol{\Delta}_{i,j}=\sigma_R^2\left(\boldsymbol{I}_{M_R}\otimes(\boldsymbol{h}_{i,j}\boldsymbol{h}_{i,j}^T)\right).$$

The minimal-rate maximization is constrained via the total available power P_R at the relay, viz.

$$\mathbb{E}\{\|\widetilde{\boldsymbol{r}}\|_2^2\}=\mathrm{Tr}\{\mathbb{E}\{\boldsymbol{W}\boldsymbol{r}\boldsymbol{r}^H\boldsymbol{W}^H\}\}$$
$$=\sum_{j=1}^{L}\sum_{i=1}^{2}p_{i,j}\|\boldsymbol{W}\boldsymbol{h}_{i,j}\|_2^2+\sigma_R^2\|\boldsymbol{W}\|_F^2\le P_R\tag{14}$$

which can be expressed with respect to (w.r.t.) $\boldsymbol{w}$ as $\boldsymbol{w}^H\boldsymbol{C}\boldsymbol{w} \le P_R$ where

$$\boldsymbol{C} = \sigma_R^2 \boldsymbol{I}_{M_R^2} + \sum_{j=1}^{L}\sum_{i=1}^{2} p_{i,j}((\boldsymbol{h}_{i,j}\boldsymbol{h}_{i,j}^H)^T \otimes \boldsymbol{I}_{M_R}). \tag{15}$$

Therefore, the design problem (i.e., min-rate maximization) in MIMO AF relay networks with L operators can be cast as the following problem:

$$\begin{aligned} \max_{\boldsymbol{w}} \ & \min_{j\in[L],\, i\in[2]} \log_2\left(1 + \frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{i,j}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}(\boldsymbol{\Upsilon}_{i,j} + \boldsymbol{\Delta}_{i,j})\boldsymbol{w} + \sigma_{i,j}^2}\right) \\ \text{s. t.} \ & \boldsymbol{w}^H\boldsymbol{C}\boldsymbol{w} \le P_R. \end{aligned} \tag{16}$$

Similar to the previous example, the inequality constraint in the above problem is active at the optimal point, so we can assume that $\boldsymbol{w}^H\boldsymbol{C}\boldsymbol{w} = P_R$. Also note that $\log_2(1+\gamma_{i,j})$ is a strictly increasing monotonic function of $\gamma_{i,j}$. Therefore, we can simplify (16) by replacing $\log_2(1+\gamma_{i,j})$ with $\gamma_{i,j}$. Thus, (16) can be equivalently written in the form of $\mathcal{P}_1$.

• *Doppler-robust waveform design for active sensing:*

In the following, we describe briefly the robust waveform design formulation of [16] for enhancing the detection of moving targets whose speed is unknown at the radar transmitter. We consider a radar system with (slow-time) transmit sequence $\boldsymbol{x} \in \mathbb{C}^N$ and receive filter $\boldsymbol{w} \in \mathbb{C}^N$. The discrete-time received signal backscattered from a moving target corresponding to the range-azimuth cell under the test can be modeled as (see, e.g. [30,31]):

$$\boldsymbol{r} = \alpha_T \boldsymbol{x} \odot \boldsymbol{p}(\nu) + \boldsymbol{c} + \boldsymbol{n}, \tag{17}$$

where α_T is a complex parameter associated with backscattering effects of the target as well as propagation effects, $\boldsymbol{p}(\nu) = [1, e^{j\nu}, \ldots, e^{j(N-1)\nu}]^T$ with ν being the normalized target Doppler shift (expressed in radians), $\boldsymbol{c}$ is the N-dimensional column vector containing clutter (signal-dependent interference) samples, and $\boldsymbol{n}$ is the N-dimensional column vector of (signal-independent) interference samples. The SINR at the output of the receive filter can be formulated as

$$\mathrm{SINR}(\nu) = \frac{|\alpha_T|^2 \left|\boldsymbol{w}^H(\boldsymbol{x}\odot\boldsymbol{p}(\nu))\right|^2}{\boldsymbol{w}^H\boldsymbol{\Sigma}_{\boldsymbol{c}}(\boldsymbol{x})\boldsymbol{w} + \boldsymbol{w}^H\boldsymbol{M}\boldsymbol{w}}, \tag{18}$$

where $\boldsymbol{M} \triangleq \mathbb{E}\{\boldsymbol{n}\boldsymbol{n}^H\}$ and $\boldsymbol{\Sigma}_{\boldsymbol{c}}(\boldsymbol{x})$ is the covariance matrix of $\boldsymbol{c}$ given by Aubry et al. [30]

$$\boldsymbol{\Sigma}_{\boldsymbol{c}}(\boldsymbol{x}) = \sum_{k=0}^{N_c-1}\sum_{i=0}^{L-1} \sigma_{(k,i)}^2 \boldsymbol{J}_k \boldsymbol{\Gamma}(\boldsymbol{x},(k,i))\boldsymbol{J}_k^T \tag{19}$$

with $\sigma_{(k,i)}^2 = \mathbb{E}\left[|\alpha_{(k,i)}|^2\right]$ being the mean interfering power associated with the clutter patch located at the (k,i)th range-azimuth bin whose Doppler shift is supposed to be uniformly distributed in the interval $\Omega_c = \left(\bar{\nu}_{d_{(k,i)}} - \frac{\epsilon_{(k,i)}}{2}, \bar{\nu}_{d_{(k,i)}} + \frac{\epsilon_{(k,i)}}{2}\right)$ [31]. Herein $\boldsymbol{\Gamma}(\boldsymbol{x},(k,i)) = \mathrm{Diag}(\boldsymbol{x})\boldsymbol{\Phi}_{\epsilon_{(k,i)}}^{\bar{\nu}_{d_{(k,i)}}}\mathrm{Diag}(\boldsymbol{x})^H$ where $\boldsymbol{\Phi}_{\epsilon_{(k,i)}}^{\bar{\nu}_{d_{(k,i)}}}(l,m)$ is the covariance matrix of $\boldsymbol{p}(\nu_{d_{(k,i)}})$ [30], viz.

$$\boldsymbol{\Phi}_{\epsilon_{(k,i)}}^{\bar{\nu}_{d_{(k,i)}}}(l,m) = \begin{cases} 1 & \text{if} \quad l = m \\ \left(\frac{\sin[0.5(l-m)\epsilon_{(k,i)}]}{[0.5(l-m)\epsilon_{(k,i)}]}\right) e^{j(l-m)\bar{\nu}_{d_{(k,i)}}} & \text{if} \quad l \neq m \end{cases} \qquad (l,m) \in \{1,\ldots,N\}^2. \tag{20}$$

The problem of Doppler robust joint design of transmit sequence $\boldsymbol{x}$ and receive filter $\boldsymbol{w}$ can be cast as the following max-min optimization problem:

$$\mathcal{P}_{\mathcal{D}}: \ \max_{\boldsymbol{x},\boldsymbol{w}} \ \min_{\nu\in\Omega} \ \left\{ \frac{\left|\boldsymbol{w}^H(\boldsymbol{x}\odot\boldsymbol{p}(\nu))\right|^2}{\boldsymbol{w}^H\boldsymbol{\Sigma}_{\boldsymbol{c}}(\boldsymbol{x})\boldsymbol{w} + \boldsymbol{w}^H\boldsymbol{M}\boldsymbol{w}} \right\}, \quad \text{s. t.} \ \ \|\boldsymbol{x}\|_2^2 = e \tag{21}$$

where $\Omega = [\nu_l, \nu_u] \subseteq [-\pi, \pi]$ denotes a given interval of the target Doppler shift ν and e denotes the maximum available transmit energy. Note that for *a priori* known target Doppler shift $\tilde{\nu}$ (i.e. $\Omega = [\tilde{\nu}, \tilde{\nu}]$), the problem $\mathcal{P}_{\mathcal{D}}$ boils down to the considered problem in [30]. The reader may now observe that for any given ν, and fixed $\boldsymbol{x}$, the objective of (21) is a fractional quadratic function of $\boldsymbol{w}$. Similarly, for any given ν, and fixed $\boldsymbol{w}$, the objective of (21) can be written as a fractional quadratic function of $\boldsymbol{x}$. As a result, for any of $\boldsymbol{x}$ or $\boldsymbol{w}$, (21) can be considered as a continuous version of $\mathcal{P}_1$, with Doppler shifts taking values within a continuous interval. On the other hand, a discrete version of (21), whose Doppler shifts occur on a discrete grid, will fit exactly the formulation of $\mathcal{P}_1$.

• *Robust classification in machine learning:*

The focus of this example is the linear discriminant analysis (LDA) method used in statistics, pattern recognition and machine learning to find a linear function of the *features* that separates several classes of objects or events [32–38]. We assume T classes with the i^{th} class of data having a mean of μ_i and the same covariance matrix $\boldsymbol{\Sigma}$. Then the scatter between classes may be defined by the sample covariance of the class means [39]

$$\boldsymbol{\Sigma}' = \frac{1}{T}\sum_{i=1}^{T}(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \tag{22}$$

where $\boldsymbol{\mu}$ is the mean of the class means. The class separation in a direction $\boldsymbol{w}$ in this case will be given by

$$\theta = \frac{\boldsymbol{w}^H\boldsymbol{\Sigma}'\boldsymbol{w}}{\boldsymbol{w}^H\boldsymbol{\Sigma}\boldsymbol{w}}. \tag{23}$$

However, we note that the class parameters may be subject to change in time-varying scenarios, or may come from different agents in networked environments. Moreover, LDA can be sensitive to errors or imperfections in the input data [40]. As a result, one may need to simultaneously optimize θ for various pairs of $(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}')$ to achieve a robust classifier, which will require tackling $\mathcal{P}_1$.

1.2. Related works and contributions of the paper

Due to its vast of area of applications, $\mathcal{P}_1$ has been studied extensively in the literature, particularly when Ω denotes a total-power or per-antenna power constraint (see e.g., [2–6,8–10] and the references therein). As a result, different approaches have been proposed to solve the design problem, including those based on uplink-downlink duality [2], the Lagrangian duality [4] and quasi-convex formulations [10]. The semidefnite relaxation (SDR) [41] is, however, the most prominent approach to the type of problems resembling to $\mathcal{P}_1$.[1] In this work, we propose a novel optimization framework (which we call Grab-n-Pull, or GnP) that can efficiently tackle $\mathcal{P}_1$. Note that the proposed framework subsumes the traditional methods handling total-power or per-antenna power constraints, while also allowing for intricate signal constraints such as unimodularity or discrete-phase requirements. In addition, the proposed method appear to outperform the widely used semidefinite relaxation approach in terms of both the quality of approximate solutions and the computational cost.

The rest of this work is organized as follows. Section 2 discusses several properties of $\mathcal{P}_1$, and presents various interesting

[1] A specific formulation of SDR to tackle $\mathcal{P}_1$ is discussed in Section 5.

problem structures with strong connection to $\mathcal{P}_1$. Section 3 discusses the proposed max-min optimization framework. Section IV further studies the proposed framework and associated parameter settings. Several numerical examples are provided in Section 5. Finally, Section 6 concludes the paper.

2. Preliminaries and related problems

In order to study $\mathcal{P}_1$, the following preliminary remarks appear to be necessary:

1) The objective of $\mathcal{P}_1$ and its optima (values) are independent to a scaling of $\boldsymbol{w}$. As a result, we can readily assume that $\boldsymbol{w}$ has a given l_2-norm. More precisely, in the sequel we assume that $\|\boldsymbol{w}\|_2^2 = P$ where $P \in \mathbb{R}_+$ is fixed. Such an assumption can be used conveniently along with other signal constraints used in practice— see the discussion on signal constraints below (42).
2) The objective of $\mathcal{P}_1$ is upper bounded via the generalized eigenvalue bound [42],

$$\frac{\boldsymbol{w}^H\boldsymbol{A}_i\boldsymbol{w}}{\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w}} \leq \sigma_{\max}\left\{\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\right\}, \quad \forall\ \boldsymbol{w} \neq 0. \tag{24}$$

Due to the *max-min inequality* [43], the latter bound implies

$$\begin{aligned} &\max_{\boldsymbol{w}}\left\{\min_{i\in[K]}\left\{\frac{\boldsymbol{w}^H\boldsymbol{A}_i\boldsymbol{w}}{\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w}}\right\}\right\} \\ &\leq \min_{i\in[K]}\left\{\max_{\boldsymbol{w}}\left\{\frac{\boldsymbol{w}^H\boldsymbol{A}_i\boldsymbol{w}}{\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w}}\right\}\right\} \\ &\leq \min_{i\in[K]}\left\{\sigma_{\max}\left\{\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\right\}\right\}. \end{aligned} \tag{25}$$

As a consequence of (25), any optimization approach that can yield a monotonically increasing sequence of the objective of $\mathcal{P}_1$ is *convergent* [44].

There are several interesting problems that have strong connections to $\mathcal{P}_1$; and thus tackling $\mathcal{P}_1$ may hold the key to approaching them. We itemize such problems below.

- Fractional quadratic programming:

$$\begin{aligned} \mathcal{P}_{1,single}:\ \max_{\boldsymbol{w}}\ & \frac{\boldsymbol{w}^H\boldsymbol{A}_i\boldsymbol{w}}{\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w}} \\ \text{s. t.}\ & \boldsymbol{w} \in \Omega, \end{aligned} \tag{26}$$

Clearly, this is a special case of $\mathcal{P}_1$ with $K = 1$.
- Max-min quadratic programs:

$$\begin{aligned} \mathcal{P}_{1,QP}:\ \max_{\boldsymbol{w}}\ & \min_{i\in[K]}\left\{\boldsymbol{w}^H\boldsymbol{A}_i\boldsymbol{w}\right\} \\ \text{s. t.}\ & \boldsymbol{w} \in \Omega, \end{aligned} \tag{27}$$

which is a special case of $\mathcal{P}_1$ with $\{\boldsymbol{B}_i\}$ set to an scaled version of the identity matrix.
- Min-max fractional quadratic programming: The proposed algorithm in this paper can be used to approach the min-max alternatives of $\mathcal{P}_1$:

$$\begin{aligned} \mathcal{P}_{1,\text{minmax}}:\ \min_{\boldsymbol{w}}\ & \max_{i\in[K]}\left\{\frac{\boldsymbol{w}^H\boldsymbol{A}_i\boldsymbol{w}}{\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w}}\right\} \\ \text{s. t.}\ & \boldsymbol{w} \in \Omega, \end{aligned} \tag{28}$$

as such problems can be easily rewritten in the form of $\mathcal{P}_1$. Namely, (28) is equivalent to

$$\begin{aligned} \max_{\boldsymbol{w}}\ & \min_{i\in[K]}\left\{\frac{\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w}}{\boldsymbol{w}^H\boldsymbol{A}_i\boldsymbol{w}}\right\} \\ \text{s. t.}\ & \boldsymbol{w} \in \Omega. \end{aligned} \tag{29}$$

- Matrix $\mathcal{P}_1$: One can tackle the matrix optimization problem

$$\begin{aligned} \mathcal{P}_{1,\text{Tr}}:\ \max_{\boldsymbol{W}}\ & \min_{i\in[K]}\left\{\frac{\text{Tr}\left(\boldsymbol{W}^H\boldsymbol{A}_i\boldsymbol{W}\right)}{\text{Tr}\left(\boldsymbol{W}^H\boldsymbol{B}_i\boldsymbol{W}\right)}\right\} \\ \text{s. t.}\ & \boldsymbol{W} \in \boldsymbol{\Omega} \end{aligned} \tag{30}$$

through $\mathcal{P}_1$ by employing the identity $\text{Tr}\left(\boldsymbol{W}^H\boldsymbol{X}\boldsymbol{W}\right) = \boldsymbol{w}^H(\boldsymbol{I}\otimes\boldsymbol{X})\boldsymbol{w}$, in which $\boldsymbol{w} = vec(\boldsymbol{W})$.

3. The max-min optimization framework

We begin by considering a reformulated version of $\mathcal{P}_1$; namely,

$$\begin{aligned} \mathcal{P}_2:\ \max_{\boldsymbol{w}}\ & \min_{i\in[K]}\{\lambda_i\} \\ \text{s. t.}\ & \boldsymbol{w} \in \Omega, \end{aligned} \tag{31}$$

$$\lambda_i = \frac{\boldsymbol{w}^H\boldsymbol{A}_i\boldsymbol{w}}{\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w}},\ \forall\, i \in [K]. \tag{32}$$

Note that (32) holds if and only if $\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\|_2^2 = \lambda_i\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2^2$, or equivalently $\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\|_2 = \sqrt{\lambda_i}\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2$. In particular, the left-hand side of (32) is *close* to the right-hand side of (32) if and only if $\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\|_2$ is *close* to $\sqrt{\lambda_i}\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2$. Therefore, by employing the auxiliary variables $\{\lambda_i\}$, one can consider the following optimization problem as an alternative to $\mathcal{P}_2$ (and $\mathcal{P}_1$):

$$\begin{aligned} \mathcal{P}_3:\ \max_{\boldsymbol{w},\{\lambda_i\}}\ & \min_{i\in[K]}\{\lambda_i\} - \eta\sum_{i=1}^{K}(\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\|_2 - \sqrt{\lambda_i}\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2)^2 \\ \text{s. t.}\ & \boldsymbol{w} \in \Omega;\ \lambda_i \geq 0,\ \forall\, i \in [K]; \end{aligned} \tag{33}$$

in which $\eta > 0$ determines the weight of the penalty-term added to the original objective of $\mathcal{P}_2$; and where $\mathcal{P}_3$ and $\mathcal{P}_2$ coincide as $\eta \to +\infty$. Note that optimizing $\mathcal{P}_3$ w. r. t. $\boldsymbol{w}$ may require rewriting $\mathcal{P}_3$ as a quartic objective in $\boldsymbol{w}$. To circumvent this, we continue by introducing $\mathcal{P}_4$—yet another alternative objective:

$$\begin{aligned} \mathcal{P}_4:\ \max_{\boldsymbol{w},\{\lambda_i\},\{\boldsymbol{Q}_i\}}\ & \min_{i\in[K]}\{\lambda_i\} - \eta\sum_{i=1}^{K}\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w} - \sqrt{\lambda_i}\boldsymbol{Q}_i\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2^2 \\ \text{s. t.}\ & \boldsymbol{w} \in \Omega,\ \lambda_i \geq 0,\ \forall\, i \in [K]; \end{aligned} \tag{34}$$

$$\boldsymbol{Q}_i^H\boldsymbol{Q}_i = \boldsymbol{I}_N,\ \forall\, i \in [K]. \tag{35}$$

To see why $\mathcal{P}_4$ and $\mathcal{P}_3$ are equivalent, observe that the minimizer $\boldsymbol{Q}_i$ of $\mathcal{P}_4$ (satisfying (35), also known as Steifel manifold [45–47]) is a unitary *rotation* matrix that aligns the vector $\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}$ in the same *direction* as $\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}$, without changing its ℓ_2-norm. More precisely, at the minimizer $\boldsymbol{Q}_i$ of $\mathcal{P}_4$, we have that

$$\boldsymbol{Q}_i\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w} = \left(\frac{\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}}{\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\|_2}\right)\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2. \tag{36}$$

Using (36), it is straightforward to verify that

$$\begin{aligned} &\sum_{i=1}^{K}\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w} - \sqrt{\lambda_i}\boldsymbol{Q}_i\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2^2 \\ &= \sum_{i=1}^{K}\left\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w} - \sqrt{\lambda_i}\left(\frac{\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}}{\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\|_2}\right)\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2\right\|_2^2 \\ &= \sum_{i=1}^{K}\left(\left\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\right\|_2 - \sqrt{\lambda_i}\left\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\right\|_2\right)^2 \end{aligned} \tag{37}$$

which concludes the observation.

In contrast to $\mathcal{P}_3$, the optimization problem $\mathcal{P}_4$ can be easily rewritten as a quadratic program (QP) in $\boldsymbol{w}$; a widely studied type of program that facilitates the usage of power method-like iterations, and thus employing different signal constraints Ω—more on this later. Note that, until now, we have shown that,

- $\mathcal{P}_1$ and $\mathcal{P}_2$ are equivalent.
- $\mathcal{P}_3$ and $\mathcal{P}_4$ are equivalent.
- $\mathcal{P}_3$ and $\mathcal{P}_4$ can be used as alternatives to the original problem, i.e. $\mathcal{P}_1$.

In the following, our goal is to

- propose an efficient iterative optimization framework based on a separate optimization of the objective of $\mathcal{P}_4$ over its three partition of variables, viz. $\boldsymbol{w}$, $\{\boldsymbol{Q}_i\}$, and $\{\lambda_i\}$, and in particular,
- study the properties of $\mathcal{P}_4$ to pave the way for an effective usage of our proposed framework in tackling fractional quadratic programs.

We note that considering $\mathcal{P}_3$ can also be useful in such a study, as $\mathcal{P}_3$ may be viewed as a simplified version of $\mathcal{P}_4$, in which the objective is already optimized w. r. t. $\{\boldsymbol{Q}_i\}$.

3.1. Power method-Like iterations (optimization w. r. t. $\mathbf{w}$)

For fixed $\{\boldsymbol{Q}_i\}$ and $\{\lambda_i\}$, one can optimize $\mathcal{P}_4$ w. r. t. $\boldsymbol{w}$ via minimizing the criterion:

$$\sum_{i=1}^{K} \|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w} - \sqrt{\lambda_i}\boldsymbol{Q}_i\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2^2 = \boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}, \tag{38}$$

where

$$\boldsymbol{R} = \sum_{i=1}^{K}\left\{(\boldsymbol{A}_i + \lambda_i\boldsymbol{B}_i) - \sqrt{\lambda_i}(\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{Q}_i\boldsymbol{B}_i^{\frac{1}{2}} + \boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{Q}_i^H\boldsymbol{A}_i^{\frac{1}{2}})\right\}. \tag{39}$$

Due to the fact that Ω enforces a fixed ℓ_2-norm on $\boldsymbol{w}$ (i.e. $\|\boldsymbol{w}\|_2^2 = P$), by defining $\hat{\boldsymbol{R}} \triangleq \mu\boldsymbol{I} - \boldsymbol{R}$ (in which $\mu > 0$ is larger than the maximum eigenvalue of $\boldsymbol{R}$), we have that

$$\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w} = -\boldsymbol{w}^H\hat{\boldsymbol{R}}\boldsymbol{w} + \underbrace{\mu P}_{const.} \tag{40}$$

Consequently, one can minimize (or decrease monotonically) the criterion in (38) by maximizing (or increasing monotonically) the objective of the following optimization problem:

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^H\hat{\boldsymbol{R}}\boldsymbol{w} \\ \text{s. t.} \quad & \boldsymbol{w} \in \Omega. \end{aligned} \tag{41}$$

Although (41) is NP-hard for a general signal constraint set [48,49], a monotonically increasing objective of (41) can be obtained using *power method-like* iterations developed in [49,50]; namely, we update $\boldsymbol{w}$ iteratively by solving the following *nearest-vector* problem at each iteration:

$$\begin{aligned} \min_{\boldsymbol{w}^{(s+1)}} \quad & \left\|\boldsymbol{w}^{(s+1)} - \hat{\boldsymbol{R}}\boldsymbol{w}^{(s)}\right\|_2 \\ \text{s. t.} \quad & \boldsymbol{w}^{(s+1)} \in \Omega, \end{aligned} \tag{42}$$

where s denotes the internal iteration number, and $\boldsymbol{w}^{(0)}$ is the current value of $\boldsymbol{w}$. Note that we can continue updating $\boldsymbol{w}$ until convergence in the objective of (41), or for a fixed number of steps, say S. A proof of the *monotonic* behavior of power method-like iterations is presented in Appendix A. The interested readers can also find an approximation method to solve (41) and (42) in [52,53].

Now, we take a deeper look at various signal constraints Ω typically used in practice, as well as their associated constrained solutions to (42):

• *Total-power constraint:* Note that the energy of designed signals should always be upper bounded in practice, which can be formulated as a total-power constraint, viz.

$$\Omega = \{\boldsymbol{w} : \|\boldsymbol{w}\|_2^2 = P\}, \quad P > 0. \tag{43}$$

In this case, the set of power method-like iterations in (42) boils down to a typical power method aiming to find the dominant eigenvector of $\hat{\boldsymbol{R}}$, however with an additional scaling to attain a power of P.

• *Per-antenna power constraint:* Power management per antenna avoids an uneven (and most likely hazardous) distribution of power over the antenna array, and is shown to be more effective than total-power constraint in some applications; see e.g. [54]. We consider K antennas each with a power of P_{ant}, and assume that $M = N/K$ entries of $\boldsymbol{w}$ are devoted to each antenna. As a result, we can solve (42) by considering the nearest-vector problem for sub-vectors associated with each antenna separately—i.e., K nearest-vector problems all with vector arguments of length M.

• *Unimodular signal design:* Unimodular codes are widely used in many radar and communication applications due to their low peak-to-average-power ratio [49,55]. The set of unimodular codes is defined as

$$\Omega = \left\{e^{j\varphi} : \varphi \in [0, 2\pi)\right\}^N. \tag{44}$$

Moreover, the unimodular solution to (42) is simply given by

$$\boldsymbol{w}^{(s+1)} = \exp\left(j\arg\left(\hat{\boldsymbol{R}}\boldsymbol{w}^{(s)}\right)\right). \tag{45}$$

• *Discrete-phase signal design:* Such signals share the low peak-to-average-power ratio property of unimodular signals, and at the same time, offer a reduced implementation complexity due to their discrete/finite nature [55,56]. We define the set of discrete-phase signals as

$$\Omega = \left\{e^{j\frac{2\pi}{Q}q} : \ q = 0, 1, \cdots, Q-1\right\}^N \tag{46}$$

where Q denotes the phase *quantization* level. The discrete-phase solution to (42) is given by

$$\boldsymbol{w}^{(s+1)} = \exp\left(j\mu_Q\left(\arg\left(\hat{\boldsymbol{R}}\boldsymbol{w}^{(s)}\right)\right)\right) \tag{47}$$

where $\mu_Q(.)$ yields (for each entry of the vector argument) the closest element in the Q-ary alphabet described in (46).

We refer the interested reader to find more details on the properties of power method-like iterations in [49–51].

3.2. Rotation-Aided fitting (optimization w. r. t. $\{\mathbf{Q}_i\}$)

Suppose $\boldsymbol{w}$ and $\{\lambda_i\}$ are fixed. As discussed earlier, the maximizer $\boldsymbol{Q}_i$ of $\mathcal{P}_4$ is a rotation matrix that maps $\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}$ in the same direction as $\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}$. Let

$$\begin{cases} \boldsymbol{u}_i = \boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w} / \|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\|_2, \\ \boldsymbol{v}_i = \boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w} / \|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2, \end{cases} \tag{48}$$

and note that (36) can be rewritten as,

$$\boldsymbol{u}_i = \boldsymbol{Q}_i\boldsymbol{v}_i \tag{49}$$

for all $i \in [K]$. We define unitary matrices $\boldsymbol{Q}_{\boldsymbol{u}_i}$ and $\boldsymbol{Q}_{\boldsymbol{v}_i}$ in $\mathbb{C}^{N\times N}$ as

$$\boldsymbol{Q}_{\boldsymbol{u}_i} = [\widehat{\boldsymbol{u}}_{i1}, \widehat{\boldsymbol{u}}_{i2}, \ldots, \widehat{\boldsymbol{u}}_{iN}] \tag{50}$$

$$\boldsymbol{Q}_{\boldsymbol{v}_i} = [\widehat{\boldsymbol{v}}_{i1}, \widehat{\boldsymbol{v}}_{i2}, \ldots, \widehat{\boldsymbol{v}}_{iN}], \tag{51}$$

where each of the sets $\{\widehat{\boldsymbol{u}}_{ij}\}_j$ and $\{\widehat{\boldsymbol{v}}_{ij}\}_j$, $j \in [N]$, builds an orthonormal basis for $\mathbb{C}^N$, and

$$\begin{cases} \widehat{\boldsymbol{u}}_{i1} = \boldsymbol{u}_i, \\ \widehat{\boldsymbol{v}}_{i1} = \boldsymbol{v}_i. \end{cases} \tag{52}$$

Table 2
Recursive Grab-n-Pull*procedure to determine* Υ.

Step 0: Set $\Upsilon = \varnothing$.
Step 1: Include 1 in Υ.
Remark: Based on Lemma 2, the primitive index 1 belongs to Υ, *as* $\sqrt{\lambda^\star}$ *is always larger than* γ_1.
Step 2: Given the current index set of minimal variables Υ, obtain $\sqrt{\lambda^\star}$ using (57).
Remark: Note that if $\sqrt{\lambda^\star}$ *is smaller than* γ_k *for all* $k \in [K]\backslash\Upsilon$ *then the obtained* Υ *is optimal, as all* λ_k *with* $k \in [K]\backslash\Upsilon$ *have chosen their values freely to maximize the objective of* Λ; *as a result, adding other indices to* Υ *will lead to a decreased objective of* Λ.
Step 3: Let $\{h\} \subset [K]$ denote the indices for which $h \notin \Upsilon$. If $\gamma_h \leq \sqrt{\lambda^\star}$, include h in Υ, and goto Step 2; otherwise stop.
Remark: This is a direct consequence of Lemma 2, particularly considering that $\sqrt{\lambda^\star}$ *is only increasing with growing* $|\Upsilon|$, *which corresponds to adding larger* γ_i*s to the weighted sum in (58).*

Based on the above definitions, the following lemma constructs the optimal $\boldsymbol{Q}_i$ (the proof of Lemma 1 is straightforward and omitted herein):

Lemma 1. *The maximizer* $\{\boldsymbol{Q}_i\}$ *of* $\mathcal{P}_4$ *can be found as*

$$\boldsymbol{Q}_i = \boldsymbol{Q}_{\boldsymbol{u}_i} \boldsymbol{Q}_{\boldsymbol{v}_i}^H \tag{53}$$

for all $i \in [K]$.

Remark 1. Note that $\{\boldsymbol{Q}_{\boldsymbol{u}_i}\}$ and $\{\boldsymbol{Q}_{\boldsymbol{v}_i}\}$ are not generally unique since $\{\widehat{\boldsymbol{u}}_{ij}\}$ and $\{\widehat{\boldsymbol{v}}_{ij}\}$ can be chosen rather arbitrarily for $j \neq 1$. This further implies that the maximizer $\{\boldsymbol{Q}_i\}$ of $\mathcal{P}_4$ is also not unique, which is in agreement with the common understanding that rotation matrices are not necessarily unique when the dimension grows large.

3.3. Grab-n-Pull (optimization w. r. t. $\{\lambda_i\}$)

Note that according to (37), once the optimal $\{\boldsymbol{Q}_i\}$ are used, the objectives of $\mathcal{P}_3$ and $\mathcal{P}_4$ can be considered interchangeably. We assume that optimal $\boldsymbol{w}$ and $\{\boldsymbol{Q}_i\}$ are obtained according to the guidelines described in Sections 3.1 and 3.2, respectively, and are fixed. Therefore, to find $\{\lambda_i\}$, we can equivalently focus on obtaining the maximizer $\{\lambda_i\}$ of $\mathcal{P}_3$ via the optimization problem:

$$\begin{aligned} \Lambda: \quad & \max_{\{\lambda_i\}} \ \min_{i\in[K]}\{\lambda_i\} - \eta \sum_{i=1}^{K} (\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\|_2 - \sqrt{\lambda_i}\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2)^2 \\ & \text{s. t.} \quad \lambda_i \geq 0, \ \forall i \in [K]. \end{aligned} \tag{54}$$

While the above problem may be reformulated and solved using convex optimization tools, a more insightful approach to optimize $\{\lambda_i\}$, referred to as Grab-n-Pull, is presented below. Particularly, the proposed approach sheds light on the behavior of λ_is in their optimal setting.

Definition 1. Let $\{\lambda_i^\star\}$ denote the optimal $\{\lambda_i\}$ of Λ, and $\lambda^\star \triangleq \min_{i\in[K]}\{\lambda_i^\star\}$. We let Υ to denote the set of all indices m for which $\lambda_m^\star$ takes the minimal value among all $\{\lambda_i^\star\}$, i.e.

$$\Upsilon = \{m \in [K] \ : \ \lambda_m^\star = \lambda^\star\}. \tag{55}$$

Moreover, we refer to $\gamma_i^2 \triangleq \|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\|_2^2 / \|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\|_2^2$ as the *shadow value* of $\lambda_i^\star$, for all $i \in [K]$.

It is straightforward to verify from the objective of (54) that if $\lambda_i^\star > \lambda^\star$, then $\lambda_i^\star = \gamma_i^2$. On the other hand, to obtain $\lambda^\star$, we need to maximize the criterion:

$$f(\lambda) = \lambda - \eta \sum_{k\in\Upsilon} \left(\|\boldsymbol{A}_k^{\frac{1}{2}}\boldsymbol{w}\|_2 - \sqrt{\lambda}\, \|\boldsymbol{B}_k^{\frac{1}{2}}\boldsymbol{w}\|_2 \right)^2. \tag{56}$$

Provided that η is large *enough* (see Section 4), the optimal $\lambda^\star$ of the quadratic criterion in (56) is given by

$$\sqrt{\lambda^\star} = \frac{\eta \sum_{k\in\Upsilon} \alpha_k \beta_k}{\eta \sum_{k\in\Upsilon} \beta_k^2 - 1} \tag{57}$$

in which $\alpha_k \triangleq \|\boldsymbol{A}_k^{\frac{1}{2}}\boldsymbol{w}\|_2$, and $\beta_k \triangleq \|\boldsymbol{B}_k^{\frac{1}{2}}\boldsymbol{w}\|_2$. It is interesting to have some insight into what $\sqrt{\lambda^\star}$ represents: Note that (57) can be rewritten as

$$\sqrt{\lambda^\star} = \frac{\sum_{k\in\Upsilon} \gamma_k \beta_k^2}{\sum_{k\in\Upsilon} \beta_k^2 - 1/\eta}. \tag{58}$$

As a result, $\sqrt{\lambda^\star}$ can be viewed as a weighted average of γ_k for $k \in \Upsilon$—except that the term $-1/\eta$ in the denominator of (58) makes $\sqrt{\lambda^\star}$ *a bit* larger than the actual weighted average. However, for an increasing η, $\sqrt{\lambda^\star}$ converges to the exact value of the weighted average specified above.

Hereafter, we propose a recursive *Grab-n-Pull* procedure to fully determine Υ. Note that once Υ is given, one can obtain $\sqrt{\lambda^\star}$ via (57). The proposed approach will make use of the following observation:

Lemma 2. *If* $\gamma_i^2 < \lambda^\star$ *for any* $i \in [K]$, *then* $i \in \Upsilon$.

Proof. The inequality $\gamma_i^2 < \lambda^\star$ implies that $\lambda_i^\star \neq \gamma_i^2$. Considering the discussion above (56), one can conclude that $\lambda_i^\star \leq \lambda^\star$, which due to the definition of $\lambda^\star$ yields $\lambda_i^\star = \lambda^\star$. Hence, the proof is complete. □

Without loss of generality, and for the sake of simplicity, we assume in the sequel that the matrix pairs $\{(\boldsymbol{A}_i, \boldsymbol{B}_i)\}$ are sorted in such a way to form the ascending order:

$$0 \leq \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K. \tag{59}$$

Based on the above reordering, the proposed approach is described in Table 2. Moreover, an illustration of the method is depicted in Fig. 1. The name of the method, i.e. *Grab-n-Pull*, comes from the intuition that the method grabs and pulls the lowest values of $\{\lambda_i\}$ to a level which is suitable for optimization of the alternative objectives, while achieving *equality*, at least for the *lowest* λ_is.

Finally, our optimization framework based on maximizing the objective of $\mathcal{P}_4$ over $\boldsymbol{w}$, $\{\boldsymbol{Q}_i\}$, and $\{\lambda_i\}$ is summarized using a flowchart in Fig. 2. Note that, due to the key role of Grab-n-Pull procedure in the proposed optimization framework, we also use the term Grab-n-Pull (or its abbreviated form GnP) when referring to the general framework. In the following section, we study different criteria in choosing a suitable η, as well as the various interesting aspects tied to the proposed framework.

4. Grab-n-Pull: convergence, settings and discussions

To perform a suitable selection of η, one should note that unlike the objective of the original problem $\mathcal{P}_1$, choosing η may be sensitive not only to $\{\boldsymbol{A}_k\}$ and $\{\boldsymbol{B}_k\}$, but also to the power, or a scaling of the signal $\boldsymbol{w}$. This can be observed easily from the penalty terms in $\mathcal{P}_3$ and $\mathcal{P}_4$ where a scaling of $\boldsymbol{w}$ can be fully compensated via a corresponding scaling in η.

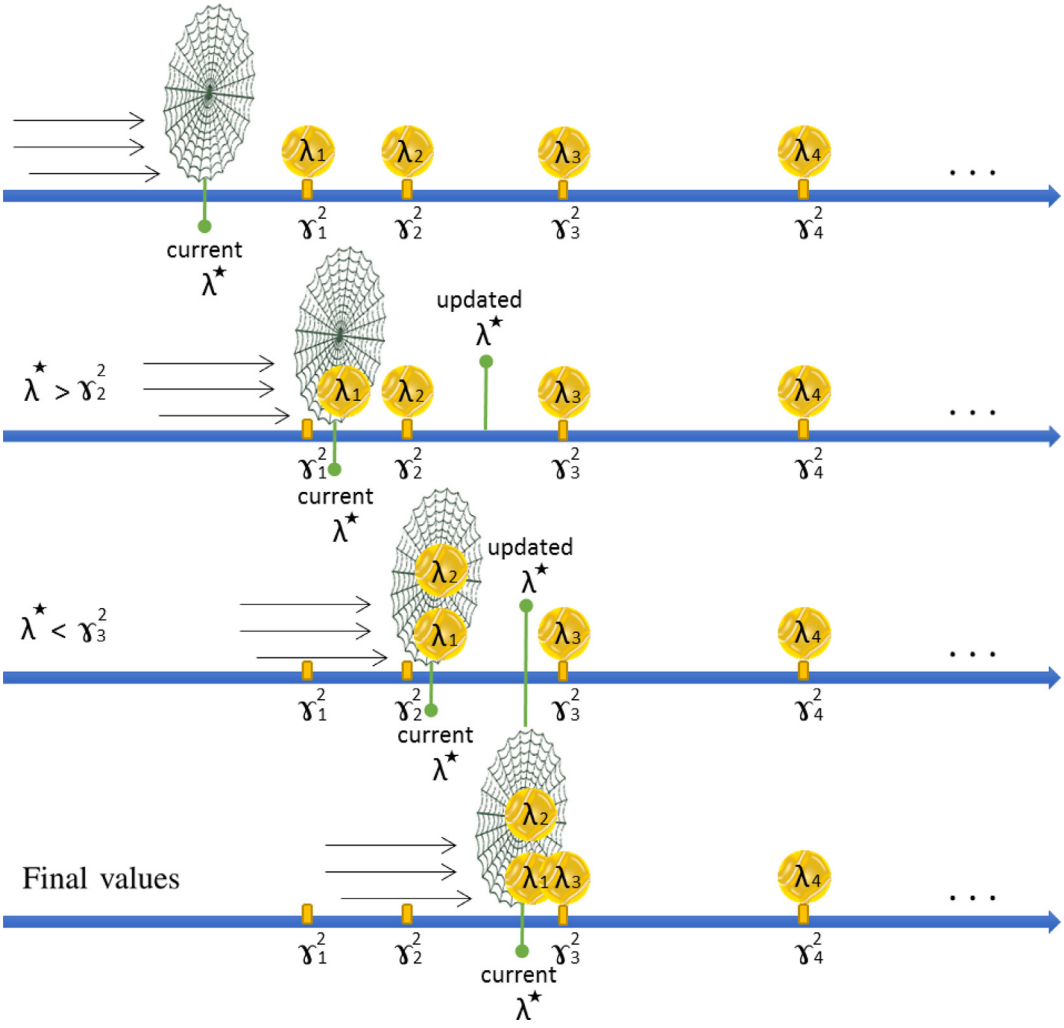

Fig. 1. An illustration of the Grab-n-Pull procedure. The optimal values of $\{\lambda_i\}$ are obtained when $\lambda^\star < \gamma_3^2$, which sets Υ to {1, 2}.

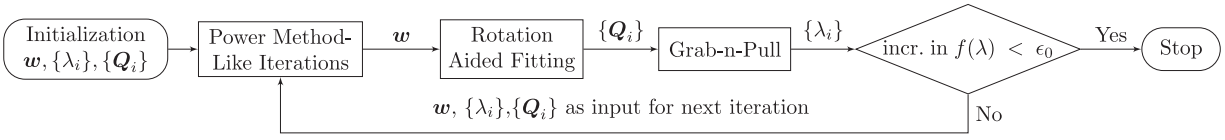

Fig. 2. Flowchart of the proposed algorithm performing the optimization w. r. t. all variables at each iteration.

4.1. Guaranteed convergence by bounding η

We begin our study from $f(\lambda)$ in (56). In particular, by defining α_k and β_k as in (57), $f(\lambda)$ can be written as

$$f(\lambda) = \lambda - \eta \sum_{k\in\Upsilon} \left(\alpha_k^2 + \lambda\beta_k^2 - 2\sqrt{\lambda}\alpha_k\beta_k\right) = \lambda\left(1 - \eta\sum_{k\in\Upsilon}\beta_k^2\right) + 2\sqrt{\lambda}\eta\left(\sum_{k\in\Upsilon}\alpha_k\beta_k\right) - \eta\sum_{k\in\Upsilon}\alpha_k^2. \quad (60)$$

Note that the above quadratic function of $\sqrt{\lambda}$ can be meaningfully maximized (with a bounded solution) if and only if $1 - \eta\sum_{k\in\Upsilon}\beta_k^2 < 0$, or equivalently,

$$\eta > \left(\sum_{k\in\Upsilon}\beta_k^2\right)^{-1} = \left(\boldsymbol{w}^H\left(\sum_{k\in\Upsilon}\boldsymbol{B}_k\right)\boldsymbol{w}\right)^{-1}. \quad (61)$$

In order to ensure the satisfaction of (61), one can choose the following conservative lower bound for η:

$$\eta > \eta_{lb} \triangleq \frac{1}{P}\left(\max_{i\in[K]}\left\{\sigma_{\min}^{-1}(\boldsymbol{B}_i)\right\}\right). \quad (62)$$

It should be emphasized that satisfying (62) also guarantees the convergence of our optimization framework, due to the following result:

Theorem 1. *Suppose η is large enough to satisfy (62). Then, $f(\lambda)$ is upper bounded (at its maximizer $\lambda^\star$, see (57)) as*

$$f(\lambda^\star) \le \frac{\sum_{k\in\Upsilon}\alpha_k^2}{\sum_{k\in\Upsilon}\beta_k^2 - 1/\eta} \le \sigma_{\max}\left\{\left(\left(\sum_{k\in\Upsilon}\boldsymbol{B}_k\right) - \frac{1}{\eta P}\boldsymbol{I}\right)^{-1}\left(\sum_{k\in\Upsilon}\boldsymbol{A}_k\right)\right\}. \quad (63)$$

Proof. Please refer to Appendix B.

Proof of Convergence: To see why Theorem 1 implies the convergence of our algorithm, observe that different steps of the proposed framework lead to an increasing objective of $\mathcal{P}_4$ (and $\mathcal{P}_3$). To guarantee convergence in terms of the objective value, we only need to show that the objective is bounded from above—a condition which will be met by satisfying (62). Thus, the proof is complete. □

4.2. Error-bound derivation

Considering the value of η as in (62) lays the ground for unfolding further connections between $\mathcal{P}_1/\mathcal{P}_2$ and $\mathcal{P}_3/\mathcal{P}_4$, as well as what the alternative objectives in $\mathcal{P}_3/\mathcal{P}_4$ represent. In particular, we show in the following that the error induced by transition to the alternating objectives $\mathcal{P}_3/\mathcal{P}_4$ (namely employing $\{\sqrt{\lambda_i}\}$ in lieu of $\{\gamma_i\}$) is always bounded by the amount of the penalty function:

Theorem 2. *Suppose η satisfies (62). In such a case, the deviation in the values of $\{\sqrt{\lambda_i}\}$ with regard to the shadow values $\{\gamma_i\}$ will be bounded as*[2]

$$\left|\min_{i\in[K]}\{\gamma_i\}-\min_{i\in[K]}\left\{\sqrt{\lambda_i}\right\}\right|^2<\varrho \tag{64}$$

where ϱ denotes the penalty function of $\mathcal{P}_3/\mathcal{P}_4$, viz.

$$\varrho=\eta\sum_{i=1}^{K}\left(\left\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\right\|_2-\sqrt{\lambda_i}\left\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\right\|_2\right)^2. \tag{65}$$

Proof. Let ε denote the l_2-norm of the *representation error*,

$$\varepsilon^2\triangleq\sum_{i=1}^{K}\left(\gamma_i-\sqrt{\lambda_i}\right)^2 \tag{66}$$

caused by using $\{\sqrt{\lambda_i}\}$ in lieu of $\{\gamma_i\}$ in the original max-min objective of $\mathcal{P}_1$. We make use of the following lemma whose proof is provided in Appendix C.

Lemma 3. *The difference of minimal $\{\gamma_i\}$ and $\{\sqrt{\lambda_i}\}$ is bounded by the representation error as*

$$\left|\min_{i\in[K]}\{\gamma_i\}-\min_{i\in[K]}\left\{\sqrt{\lambda_i}\right\}\right|^2\le\varepsilon^2. \tag{67}$$

Now it remains to verify that $\varepsilon^2\le\varrho$:

$$\begin{aligned}\varrho&=\eta\sum_{i=1}^{K}\left(\left\|\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{w}\right\|_2-\sqrt{\lambda_i}\left\|\boldsymbol{B}_i^{\frac{1}{2}}\boldsymbol{w}\right\|_2\right)^2\\&=\eta\sum_{i=1}^{K}(\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w})\left(\gamma_i-\sqrt{\lambda_i}\right)^2\\&\ge\sum_{i=1}^{K}\left(\gamma_i-\sqrt{\lambda_i}\right)^2=\varepsilon^2\end{aligned} \tag{68}$$

as $\eta(\boldsymbol{w}^H\boldsymbol{B}_i\boldsymbol{w})>1$ holds for all $i\in[K]$, due to (62). This concludes the proof. □

By mathematically bounding the error, the above result reaffirms that the smaller we can get the penalty function, the smaller the difference between the original and the alternative objectives becomes. In particular, if the penalty function attain zero then the original and alternative problems are *equivalent*. Now, let $g(\{\lambda_i\},\{\gamma_i^2\})$ denote the objective function of $\mathcal{P}_3/\mathcal{P}_4$. It follows from (68) that

$$\begin{aligned}g(\{\lambda_i\},\{\gamma_i^2\})&=\min_{i\in[K]}\{\lambda_i\}-\varrho\\&\le\min_{i\in[K]}\{\lambda_i\}-\varepsilon^2\triangleq\tilde{g}(\{\lambda_i\},\{\gamma_i^2\}).\end{aligned} \tag{69}$$

Note that while $\tilde{g}$ shares the same first term (i.e. $\min_{i\in[K]}\{\lambda_i\}$) as g, its second term denotes the representation error. In other words, $\tilde{g}$ replaces the original objective $\min_{i\in[K]}\{\gamma_i\}$ with $\min_{i\in[K]}\{\lambda_i\}$, and considers a penalty function that clearly represents the error induced by such change of variables. From this point of view, $\tilde{g}$ can be a very interesting objective to look at instead of the objective of $\mathcal{P}_1/\mathcal{P}_2$. The fascinating fact here is that by maximizing g (which is an easier task to accomplish), we are also maximizing $\tilde{g}$ at the same time, as g *minorizes* $\tilde{g}$. We refer the reader to a review of minorization-maximization (MM) scheme in [57,58].

4.3. On the penalty coefficient: the larger, the better?

Although with a larger η one may expect a lower value of the penalty functions in $\mathcal{P}_3$ and $\mathcal{P}_4$, a lower η can play a useful role in speeding up the algorithm. Remember that, given the index set Υ, the maximizer of $f(\lambda)$ is given by

$$\sqrt{\lambda^\star}=\frac{\sum_{k\in\Upsilon}\gamma_k\beta_k^2}{\sum_{k\in\Upsilon}\beta_k^2-1/\eta}. \tag{70}$$

As discussed earlier, for a finite $\eta>0$, $\sqrt{\lambda^\star}$ is larger than the below wighted sum of $\{\gamma_k\}_{k\in[K]}$:

$$\left(\sum_{k\in\Upsilon}\gamma_k\beta_k^2\right)/\left(\sum_{k\in\Upsilon}\beta_k^2\right), \tag{71}$$

which leads to the following *bootstrapping effect*.

Remark 2 (Bootstrapping Effect): For the sake of simplicity, assume Υ has a cardinality of one (including solely a generic index k), and consider the associated objective function of Λ:

$$f(\lambda)=\lambda-\eta\left(\alpha_k-\sqrt{\lambda}\beta_k\right)^2 \tag{72}$$

The goal of employing the penalty function in (72) is for λ to be as close as possible to its shadow value $\gamma_k^2=\alpha_k^2/\beta_k^2=(\boldsymbol{w}^H\boldsymbol{A}_k\boldsymbol{w})/(\boldsymbol{w}^H\boldsymbol{B}_k\boldsymbol{w})$. Note that:

(a) For a finite $\eta>0$, the maximizer $\lambda^\star$ of $f(\lambda)$ is larger than γ_k^2:

$$\lambda^\star=\left(\beta_k^2/\left(\beta_k^2-1/\eta\right)\right)\gamma_k^2. \tag{73}$$

(b) Only the penalty term of (72) is variable with $\boldsymbol{w}$. In particular, for a fixed λ, optimization w. r. t. $\boldsymbol{w}$ will be performed to achieve a γ_k^2 as close as possible to $\lambda^\star$.

(c) Then, thanks to (a), $\lambda^\star$ will be chosen to be larger than the current value of γ_k^2, and the same phenomenon persists by continuing with (b).

In sum, *an increased γ_k^2 will lead to an increased $\lambda^\star$, and an increased $\lambda^\star$ will lead to an increased γ_k^2, until convergence.* It is worth noting that a similar behavior can be observed for $|\Upsilon|>1$.

Now note that while the bootstrapping effect occurs for any finite $\eta>0$ satisfying (62), the penalty coefficient η can be viewed as a tuning parameter for the speed of the algorithm. Specifically, one can easily see that if η is large, $\lambda^\star$ will be slightly greater than the weighted average of $\{\gamma_k^2\}_{k\in\Upsilon}$, whereas for smaller values of η, the bounces from the weighted average are much larger—and the bootstrapping process can occur much quicker. In a related observation, one can also verify that the choice of η affects the cardinality of Υ. By evaluating $f(\lambda)$ at its maximizer $\lambda^\star$ (see (70)) we

[2] It might be useful to mention that the error bound is given for $\left(\{\gamma_i\},\{\sqrt{\lambda_i}\}\right)$, which is as effective as a bound derived for $\left(\{\gamma_i^2\},\{\lambda_i\}\right)$.

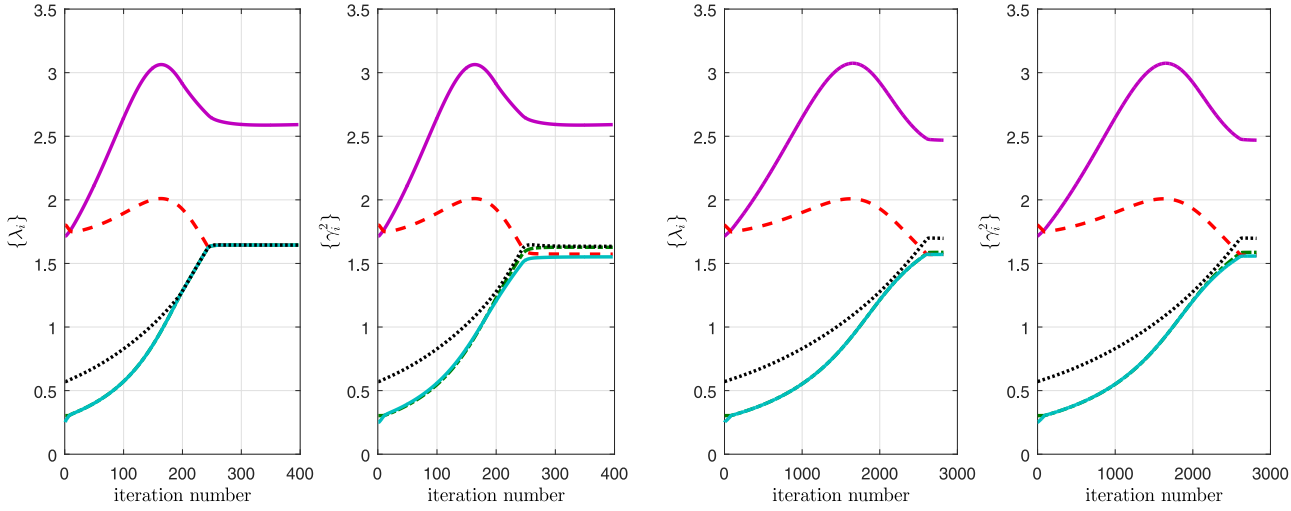

Fig. 3. Transition of the GnP optimization parameters $\{\lambda_i\}$ and $\{\gamma_i^2\}$ (distinguished by colors and line-styles) vs. the iteration number for different weights (η) of the penalty-term in $\mathcal{P}_3$ and $\mathcal{P}_4$: (a) $\eta = 1$, and (b) $\eta = 10$.

obtain

$$f(\lambda^\star) = \eta\left(\frac{\eta(\sum_{k\in\Upsilon}\alpha_k\beta_k)^2}{\eta\sum_{k\in\Upsilon}\beta_k^2 - 1} - \sum_{k\in\Upsilon}\alpha_k^2\right) \tag{74}$$

$$= \left(\left(\sum_{k\in\Upsilon}\beta_k^2\right) - 1/\eta\right)^{-1} \times \left(\sum_{k\in\Upsilon}\alpha_k^2 + \eta\left[\underbrace{\left(\sum_{k\in\Upsilon}\alpha_k\beta_k\right)^2 - \left(\sum_{k\in\Upsilon}\beta_k^2\right)\left(\sum_{k\in\Upsilon}\alpha_k^2\right)}_{(*)}\right]\right). \tag{75}$$

According to the Cauchy–Schwarz inequality, $(*)$ is always less than or equal to zero. Also, the equality is attained only if all $\{\gamma_k\}_{k\in\Upsilon}$ are identical, or $|\Upsilon| = 1$. As a result, we can conclude that for *sufficiently large* values of η, a pair $(\Upsilon, \lambda^\star)$ can serve as a solution to Λ only if $|\Upsilon|$ is kept to its minimum, given by the number of minimal $\{\gamma_k\}$ which are also identical. This implies that at the initial iterations of the method when $\{\gamma_k\}$ are most likely to be distinct, we should have $|\Upsilon| = 1$; this can cause difficulty combined with the fact that a large η requires $\lambda^\star$ to take many small steps while moving away from γ_1^2 and reaching other $\{\gamma_k\}$, and thus increasing $|\Upsilon|$ (assuming the ordering in (59)). We should add that the Grab-n-Pull procedure is most useful when $|\Upsilon| > 1$, or equivalently when η is not *very* large.

We conclude this section by discussing the trade-off originated from the selection of η, namely the question of a higher convergence speed vs. a smaller value of the penalty functions in $\mathcal{P}_3$ and $\mathcal{P}_4$. To devise a reasonable approach to this trade-off, we consider the following insight: While we can tackle $\mathcal{P}_1$ by blindly increasing all the fractions in the max-min structure, $\mathcal{P}_3$ and $\mathcal{P}_4$ suggest an alternative approach by employing the auxiliary variables $\{\lambda_i\}$. The role of $\{\lambda_i\}$ is to *determine* the *increasing* levels $\{(\boldsymbol{w}^H\boldsymbol{A}_k\boldsymbol{w})/(\boldsymbol{w}^H\boldsymbol{B}_k\boldsymbol{w})\}$ should converge to, and the optimization w. r. t. $\boldsymbol{w}$ will be performed such that $\{(\boldsymbol{w}^H\boldsymbol{A}_k\boldsymbol{w})/(\boldsymbol{w}^H\boldsymbol{B}_k\boldsymbol{w})\}$ can get close to those levels. Consequently, putting less focus on small values of the penalty functions specified above, only makes the proposed method closer to the blind approach—while leaving us with the interesting advantages of the proposed framework including the quadratic nature of the objective, efficiency, and possibility of working with various signal constraints.

5. Numerical examples

In this section, we provide several numerical examples to investigate the performance of the proposed method (performing the optimization w. r. t. all variables at each iteration). We first study the impact of choosing η on the performance of GnP. We then compare GnP with SDR in terms of run-time and quality of approximate solutions. Additionally, we provide an application-driven example, in which the GnP algorithm is used to tackle a network beamforming problem.

For numerical evaluations, random PSD matrices $\{\boldsymbol{A}_i\}$ and $\{\boldsymbol{B}_i\}$ are generated using the formula,

$$\boldsymbol{A}_i = \boldsymbol{X}_i\boldsymbol{X}_i^H, \quad \boldsymbol{B}_i = \boldsymbol{Y}_i\boldsymbol{Y}_i^H, \quad \forall\, i \in [K], \tag{76}$$

where $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$ are random matrices in $\mathbb{C}^{N\times N}$ whose elements are i.i.d. circularly symmetric zero-mean complex Gaussian random variables with variance $\sigma^2 = 1$.

The run-times are obtained on a standard PC with 4 GB memory and 2.80 GHz processor. In all examples, we stop the optimization iterations whenever the increase in the objective becomes smaller than $\epsilon_0 = 10^{-6}$. Recall from Definition 1 that at optimum $\lambda^\star = \min_{i\in[K]}\{\lambda_i\}$. Accordingly, we make use of the definition $\gamma^{\star 2} \triangleq \min_{i\in[K]}\{\gamma_i^2\}$, where $\{\gamma_i^2\}$ are the shadow values described in Definition 1.

5.1. Impact of the penalty coefficient (η)

A discussion on the impact of η was provided in Section 4.3. Herein, we present a numerical example that illustrates the outcome of choosing various values of η in connection to our previous discussions. We consider a scenario with $(K, N) = (5, 5)$ where the signal $\boldsymbol{w}$ belongs to set of unimodular vectors (defined in (44)). The matrices $\{\boldsymbol{A}_i\}$ and $\{\boldsymbol{B}_i\}$ are generated according to (76), and the optimization problem $\mathcal{P}_3$ (equivalently $\mathcal{P}_4$) is solved for different values of η.

Note that while the shadow values $\{\gamma_i^2\}$ represent the value of the fractional quadratic terms in the original objective $\mathcal{P}_1$, the auxiliary variables $\{\lambda_i\}$ tend to be *as close as possible* to $\{\gamma_i^2\}$ depending on the weight (η) of the penalty-terms in $\mathcal{P}_3$ and $\mathcal{P}_4$. In Fig. 3,

Table 3
Comparison of the performance of GnP and SDR for 100 random realization of $\{\boldsymbol{A}_i\}$ and $\{\boldsymbol{B}_i\}$, and $N = 5$.

K	η	Average $\gamma^{\star 2}/\upsilon_{\text{SDR}}$ (GR1)	Average $\gamma^{\star 2}/\upsilon_{\text{SDR}}$ (GR2)	Average $\gamma^{\star 2}/\upsilon^{\star}_{\text{SDR}}$	Average GnP CPU time (s)	Average SDR time / Average GnP time (GR1)	Average SDR time / Average GnP time (GR2)	Average number of GRs (in 1000) (GR2)
10	1	1.054	1.046	0.8941	1.7034	4.44	10.093	48.7
	0.5/10/1000	1.11	1.105	0.9456	2.4733	3.20	10.045	79.8
15	1	1.085	1.063	0.8486	2.8970	2.73	10.042	84.6
	0.5/10/1000	1.123	1.103	0.8818	3.4127	2.47	10.034	99.6
20	0.3	1.067	1.046	0.7753	2.6522	3.32	10.063	58.47
	1	1.096	1.059	0.7884	4.1367	2.11	10.038	105.9
	0.5/10/1000	1.143	1.112	0.8279	4.3395	1.96	10.034	113.1

we present the transition of variables $\{\lambda_i\}$ and $\{\gamma_i^2\}$ vs. the iteration number for two different settings of η; namely $\eta = 1$ and $\eta = 10$. As discussed earlier, although with a larger η one may expect a lower value of the penalty functions in $\mathcal{P}_3$ and $\mathcal{P}_4$, a lower η can play a useful role in speeding up the algorithm. This phenomenon can also be observed in Fig. 3, noting that the aforementioned values of η are chosen to accentuate the trade-off originated from the selection of η.

5.2. Comparison with SDR

In order to examine the performance of the proposed method, we compare it with the widely used SDR approach [41]. Considering the total power constraint on the signal ($\|\boldsymbol{w}\|_2^2 = 1$), $\mathcal{P}_1$ can be equivalently reformulated as,

$$\mathcal{R}_1 : \max_{\boldsymbol{W}} \quad \min_{i\in[K]} \left\{ \frac{\text{Tr}(\boldsymbol{A}_i\boldsymbol{W})}{\text{Tr}(\boldsymbol{B}_i\boldsymbol{W})} \right\}$$
$$\text{s. t.} \quad \text{Tr}(\boldsymbol{W}) = 1\,, \boldsymbol{W} \succeq \boldsymbol{0}\,, \text{Rank}(\boldsymbol{W}) = 1, \tag{77}$$

where $\boldsymbol{W} = \boldsymbol{w}\boldsymbol{w}^H$. Relaxing the rank-one constraint and noting that the objective function is quasi-concave, we can write the corresponding *feasibility problem* as follows:

$$\mathcal{R}_2 : \text{find} \quad \boldsymbol{W}$$
$$\text{s. t.} \quad \frac{\text{Tr}(\boldsymbol{A}_i\boldsymbol{W})}{\text{Tr}(\boldsymbol{B}_i\boldsymbol{W})} \geq \mu\,, \qquad \forall\, i \in [K],$$
$$\text{Tr}(\boldsymbol{W}) = 1\,, \boldsymbol{W} \succeq \boldsymbol{0}\,. \tag{78}$$

Note that a maximal value of μ may be found using the bisection method. Particularly, $\mathcal{R}_2$ followed by the bisection procedure is equivalent to a relaxed version of $\mathcal{R}_1$ in which the rank-one constraint is dropped. Moreover, for any given μ, $\mathcal{R}_2$ is a (convex) semidefinite program and can be solved using interior-point solvers [41]. We stop the bisection in the solver whenever the increments in μ become bounded by 10^{-6}.

Let $\boldsymbol{W}^\star$ denote the solution to $\mathcal{R}_2$ after the bisection procedure is complete. Due to the rank relaxation in $\mathcal{R}_2$, $\boldsymbol{W}^\star$ will not (in general) be rank-one. In this case, the Gaussian randomization (GR) method [41,48,59] is typically used to generate L candidates (from which the one leading to the largest objective will be chosen) for approximating the optimum $\boldsymbol{w}^\star$ of $\mathcal{P}_1$: Let $\boldsymbol{W}^\star = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^H$ be the eigen-decomposition of $\boldsymbol{W}^\star$. The lth candidate ($l \in [L]$) can be generated as $\boldsymbol{w}_l = \boldsymbol{V}\boldsymbol{\Sigma}^{1/2}\boldsymbol{v}_l$, where $\boldsymbol{v}_l \in \mathbb{C}^N \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ [8]. Note that each $\boldsymbol{w}_l$ may be scaled in order to satisfy the constraint $\|\boldsymbol{w}_l\|_2^2 = 1$. We denote the best candidate by $\boldsymbol{w}^\star_{GR}$. The corresponding objective value is thus given by

$$\upsilon_{\text{SDR}} = \min_{i\in[K]} \left\{ \frac{\boldsymbol{w}^{\star\,H}_{GR}\boldsymbol{A}_i\,\boldsymbol{w}^\star_{GR}}{\boldsymbol{w}^{\star\,H}_{GR}\boldsymbol{B}_i\,\boldsymbol{w}^\star_{GR}} \right\}. \tag{79}$$

As $\mathcal{R}_2$ followed by bisection is a relaxed version of $\mathcal{R}_1$ (and $\mathcal{P}_1$), the optimal value of its objective yields an upper bound on the optimal objectives of $\mathcal{R}_1$ (and $\mathcal{P}_1$); although it may not be tight. This upper bound is given as

$$\upsilon^\star_{\text{SDR}} = \min_{i\in[K]} \left\{ \frac{\text{Tr}(\boldsymbol{A}_i\boldsymbol{W}^\star)}{\text{Tr}(\boldsymbol{B}_i\boldsymbol{W}^\star)} \right\}. \tag{80}$$

which may be used to examine further the goodness of approximate solutions provided by various methods.

We compare the proposed method and SDR with two different settings of the GR process. In the first scenario (GR1), we limit the number of GRs to $L = 1000$. In the second scenario (GR2), we stop the GR process whenever the SDR computations (including the GR process) lasts at least 10 times of the CPU time required by GnP. This gives enough time to perform very large number of GRs, in the order of several 10000 randomizations. Table 3 presents the performance comparison of GnP and SDR averaged over 100 random realizations of $\{\boldsymbol{A}_i\}$ and $\{\boldsymbol{B}_i\}$ with $N = 5$, a set of values for K, and various η for both scenarios GR1 and GR2. Other than setting a constant value for η, inspired by the literature, we have considered the case with η being increased in two steps ($0.5 \to 100 \to 1000$) after achieving convergence for each η in use.

As expected, GR2 provides a (slightly) better performance compared to GR1, however, its run-time is considerably larger than that of GR1 thanks to the very large number of randomizations. Most importantly, it is interesting to observe that, in all cases, GnP outperforms SDR in terms of both run-time and the obtained objective value of $\mathcal{P}_1$. In other words, it appears that SDR cannot achieve the same quality of approximate solutions even with many randomizations (the GR2 scenario) which effectively makes SDR computationally exhaustive. This is presumably due to the specific fractional structure of $\mathcal{P}_1$ (in contrast to the many quadratic optimization problems for which SDR is known to yield quality results with a moderate number of randomizations [41]).

5.3. Application to multigroup multicast precoding

Finally, we will use GnP to solve the max-min network precoding problem for achieving *fairness* in a multigroup multicast scenario. A formulation of this problem in the form of $\mathcal{P}_1$ was detailed in Section 1.1. We consider a downlink transmitter with $n_{\text{Tx}} = 4$ antennas, as well as $K = 12$ single-antenna users which are divided into $G = 2$ multicast groups of 6 users. The entries of the channel vectors $\boldsymbol{h}_i$ are drawn from an i.i.d. complex Gaussian distribution with zero-mean and unit-variance. The Gaussian noise components received at each user antenna are assumed to have unit variance, i.e. $\sigma_i^2 = 1$ for all $i \in [K]$. We consider a normalized total-power constraint, i.e. with $P = 1$, and stop the optimization iterations whenever the objective increase becomes bounded by $\epsilon = 10^{-6}$.

Table 4 summarizes the results of the max-min precoding design for 300 random realizations of the multigroup multicasting channel. Average performance of the GnP for different values of η is compared with SDR followed by 1000 GRs. It can be seen that $\eta = 10$ leads to higher objective of $\mathcal{P}_1$ but increases the run-time

Table 4
Comparison of the performance of GnP and SDR in multigroup multicast network precoding for 300 random realization of the channel ($N = 8$, $K = 12$).

K	η	Average $\gamma^{\star 2}/\upsilon_{\text{SDR}}$	Average $\gamma^{\star 2}/\upsilon^{\star}_{\text{SDR}}$	Average GnP CPU time (s)	Average SDR time / Average GnP time
12	**1**	1.16	0.7796	1.4034	7.5013
	10	1.25	0.8827	8.7017	1.2139
	0.5/10/1000	1.21	0.8449	2.8303	3.6625

of the algorithm. On the other hand, increasing η in a few steps, i.e. the case with $\eta = 0.5/10/1000$, provides a good balance between the solution quality and run-time, while outperforming SDR in both criteria.

6. Concluding remarks

An optimization framework for efficient max-min fractional quadratic programming was proposed and studied. The results can be summarized as follows:

- A quadratic alternative of the original problem was proposed. Thanks to this reformulation, the proposed method can handle different signal constraints by employing the power method-like iterations. Moreover, the proposed method enjoys a low computational cost owing to the simple tasks to be performed at each iteration.
- Various aspects of the proposed approach were studied. It was shown that a lower value of the penalty coefficient η can play a useful role in speeding up the algorithm. To ensure the effectiveness of the proposed framework, a set of lower bounds on η were established.
- It was shown through numerical examples that the proposed method outperforms SDR (a widely used approach in the literature) in terms of both quality of the approximate solutions and the computational cost.

Based on the above, the proposed framework presents many unique potentials in tackling max-min fractional quadratic optimization problems, which can be of significant interest in signal and information processing applications—many of which are detailed in Section 1.1.

Appendix A. Power method-like iterations: proof of monotonicity of optimization objective

While the monotonic increase in quadratic optimization objectives when applying power method-like iterations has been established in previous works [49–51], a short proof will be provided herein for reader's convenience. Let $\boldsymbol{w}^{(s+1)}$ denote an update of the vector $\boldsymbol{w}$ provided by power method-like iterations. For a fixed $\boldsymbol{w}^{(s)}$, the said vector $\boldsymbol{w}^{(s+1)}$ is the minimizer of the criterion in (42), i.e.,

$$\left\|\boldsymbol{w}^{(s+1)} - \hat{\boldsymbol{R}}\boldsymbol{w}^{(s)}\right\|_2^2 = \text{const} - 2\,\Re\left\{\boldsymbol{w}^{(s+1)\,H}\,\hat{\boldsymbol{R}}\boldsymbol{w}^{(s)}\right\} \tag{81}$$

which implies that $\boldsymbol{w}^{(s+1)}$ yields the largest value of $\Re\left\{\boldsymbol{w}^{(s+1)\,H}\,\hat{\boldsymbol{R}}\boldsymbol{w}^{(s)}\right\}$ among all vectors in Ω. Now note that

$$\left(\boldsymbol{w}^{(s+1)} - \boldsymbol{w}^{(s)}\right)^H \hat{\boldsymbol{R}}\left(\boldsymbol{w}^{(s+1)} - \boldsymbol{w}^{(s)}\right) \geq 0. \tag{82}$$

As a result,

$$\begin{aligned}\boldsymbol{w}^{(s+1)\,H}\hat{\boldsymbol{R}}\boldsymbol{w}^{(s+1)} &\geq 2\,\Re\left\{\boldsymbol{w}^{(s+1)\,H}\,\hat{\boldsymbol{R}}\boldsymbol{w}^{(s)}\right\} - \boldsymbol{w}^{(s)\,H}\hat{\boldsymbol{R}}\boldsymbol{w}^{(s)} \\ &\geq \boldsymbol{w}^{(s)\,H}\hat{\boldsymbol{R}}\boldsymbol{w}^{(s)},\end{aligned} \tag{83}$$

due to the fact that $\Re\left\{\boldsymbol{w}^{(s+1)\,H}\,\hat{\boldsymbol{R}}\boldsymbol{w}^{(s)}\right\} \geq \boldsymbol{w}^{(s)\,H}\hat{\boldsymbol{R}}\boldsymbol{w}^{(s)}$.

Appendix B. Proof of Theorem 1

By evaluating $f(\lambda)$ at its maximizer $\lambda^\star$ (see (57)) we obtain

$$f(\lambda^\star) = \eta\left(\frac{\eta\left(\sum_{k\in\Upsilon}\alpha_k\beta_k\right)^2}{\eta\sum_{k\in\Upsilon}\beta_k^2 - 1} - \sum_{k\in\Upsilon}\alpha_k^2\right) \tag{84}$$

$$\begin{aligned}&= \left(\frac{1}{\sum_{k\in\Upsilon}\beta_k^2 - 1/\eta}\right) \\ &\times\left(\sum_{k\in\Upsilon}\alpha_k^2 + \eta\left[\left(\sum_{k\in\Upsilon}\alpha_k\beta_k\right)^2 - \left(\sum_{k\in\Upsilon}\beta_k^2\right)\left(\sum_{k\in\Upsilon}\alpha_k^2\right)\right]\right)\end{aligned} \tag{85}$$

$$\leq \frac{\sum_{k\in\Upsilon}\alpha_k^2}{\sum_{k\in\Upsilon}\beta_k^2 - 1/\eta} = \frac{\boldsymbol{w}^H\left(\sum_{k\in\Upsilon}\boldsymbol{A}_k\right)\boldsymbol{w}}{\boldsymbol{w}^H\left(\left(\sum_{k\in\Upsilon}\boldsymbol{B}_k\right) - \frac{1}{\eta P}\boldsymbol{I}\right)\boldsymbol{w}} \tag{86}$$

$$\leq \sigma_{\max}\left\{\left(\left(\sum_{k\in\Upsilon}\boldsymbol{B}_k\right) - \frac{1}{\eta P}\boldsymbol{I}\right)^{-1}\left(\sum_{k\in\Upsilon}\boldsymbol{A}_k\right)\right\} \tag{87}$$

according to the generalized eigenvalue upper bound. Note that the transition from (85) and (86) is made possible by using the Cauchy-Schwarz inequality. The proof is complete.

Appendix C. Proof of Lemma 3

Due to the defintion of ε, we have

$$\left|\gamma_i - \sqrt{\lambda_i}\right| \leq \varepsilon, \quad \forall i \in [K]. \tag{88}$$

Suppose $\min_{i\in[K]}\{\gamma_i\} \leq \min_{i\in[K]}\left\{\sqrt{\lambda_i}\right\}$, and note that according to (88) there exists λ_j ($j \in [K]$) such that

$$\left|\min_{i\in[K]}\{\gamma_i\} - \sqrt{\lambda_j}\right| < \varepsilon. \tag{89}$$

Clearly, $\min_{i\in[K]}\{\gamma_i\} \leq \min_{i\in[K]}\left\{\sqrt{\lambda_i}\right\} \leq \sqrt{\lambda_j}$ which implies

$$\left|\min_{i\in[K]}\{\gamma_i\} - \min_{i\in[K]}\left\{\sqrt{\lambda_i}\right\}\right| \leq \varepsilon. \tag{90}$$

A similar argument can be presented in the case $\min_{i\in[K]}\{\gamma_i\} > \min_{i\in[K]}\left\{\sqrt{\lambda_i}\right\}$, leading again to (90). The difference of objectives is thus contained as

$$\left|\min_{i\in[K]}\{\gamma_i\} - \min_{i\in[K]}\left\{\sqrt{\lambda_i}\right\}\right|^2 \leq \varepsilon^2. \tag{91}$$

References

[1] M. Soltanalian, A. Gharanjik, B. Shankar, B. Ottersten, Grab-n-Pull: an optimization framework for fairness-achieving networks, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016. Shanghai, China

[2] M. Schubert, H. Boche, Solution of the multiuser downlink beamforming problem with individual SINR constraints, IEEE Trans. Veh. Technol. 53 (1) (2004) 18–28.
[3] D. Cai, T. Quek, C.W. Tan, A unified analysis of max-min weighted SINR for MIMO downlink system, IEEE Trans. Signal Process. 59 (8) (2011) 3850–3862.
[4] A. Wiesel, Y. Eldar, S. Shamai, Linear precoding via conic optimization for fixed MIMO receivers, IEEE Trans. Signal Process. 54 (1) (2006) 161–176.
[5] Y.-F. Liu, Y.-H. Dai, Z.-Q. Luo, Max-min fairness linear transceiver design for a multi-user MIMO interference channel, IEEE Trans. Signal Process. 61 (9) (2013) 2413–2423.
[6] C.W. Tan, M. Chiang, R. Srikant, Maximizing sum rate and minimizing MSE on multiuser downlink: optimality, fast algorithms and equivalence via max-min SINR, IEEE Trans. Signal Process. 59 (12) (2011) 6127–6143.
[7] N.D. Sidiropoulos, T.N. Davidson, Z.-Q.T. Luo, Transmit beamforming for physical-layer multicasting, IEEE Trans. Signal Process. 54 (6) (2006) 2239–2251.
[8] E. Karipidis, N.D. Sidiropoulos, Z.-Q. Luo, Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups, IEEE Trans. Signal Process. 56 (3) (2008) 1268–1279.
[9] D.W. Cai, T.Q. Quek, C.W. Tan, S.H. Low, Max-min SINR coordinated multipoint downlink transmission—duality and algorithms, IEEE Trans. Signal Process. 60 (10) (2012) 5384–5395.
[10] G. Dartmann, X. Gong, W. Afzal, G. Ascheid, On the duality of the max-min beamforming problem with per-antenna and per-antenna-array power constraints, IEEE Trans. Veh. Technol. 62 (2) (2013) 606–619.
[11] T. Van Chien, E. Björnson, E.G. Larsson, Joint pilot sequence design and power control for max-min fairness in uplink massive MIMO, in: Proceedings of the IEEE International Conference on Communications (ICC), IEEE, 2017, pp. 1–6.
[12] A. Pascual-Iserte, D.P. Palomar, A. Pérez-Neira, M.Á. Lagunas, et al., A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization, IEEE Trans. Signal Process. 54 (1) (2006) 346–360.
[13] A.B. Gershman, Robustness Issues in Adaptive Beamforming and high-Resolution Direction Finding, New York: Marcel Dekker, 2003.
[14] S.-J. Kim, S. Boyd, A minimax theorem with applications to machine learning, signal processing, and finance, SIAM J. Optim. 19 (3) (2008) 1344–1367.
[15] A. De Maio, Y. Huang, M. Piezzo, A doppler robust max-min approach to radar code design, IEEE Trans. Signal Process. 58 (9) (2010) 4943–4947.
[16] M.M. Naghsh, M. Soltanalian, P. Stoica, M. Modarres-Hashemi, A. De Maio, A. Aubry, A doppler robust design of transmit sequence and receive filter in the presence of signal-dependent interference, IEEE Trans. Signal Process. 62 (4) (2014) 772–785.
[17] E. Bjornson, G. Zheng, M. Bengtsson, B. Ottersten, Robust monotonic optimization framework for multicell MISO systems, IEEE Trans. Signal Process. 60 (5) (2012) 2508–2523.
[18] A. Gershman, N. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, B. Ottersten, Convex optimization-based beamforming, IEEE Signal Process. Mag. 27 (3) (2010) 62–75.
[19] A. Zappone, E. Jorswieck, Energy efficiency in wireless networks via fractional programming theory, Found. Trends® Commun. Inf. Theory 11 (3–4) (2015) 185–396.
[20] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, M. Debbah, Energy-efficient power control: a look at 5G wireless technologies, IEEE Trans. Signal Process. 64 (7) (2016) 1668–1683.
[21] K.T.K. Cheung, S. Yang, L. Hanzo, Achieving maximum energy-efficiency in multi-relay OFDMA cellular networks: a fractional programming approach, IEEE Trans. Commun. 61 (7) (2013) 2746–2757.
[22] S. He, Y. Huang, S. Jin, F. Yu, L. Yang, Max-min energy efficient beamforming for multicell multiuser joint transmission systems, IEEE Commun. Lett. 17 (10) (2013) 1956–1959.
[23] J. Denis, S. Smirani, B. Diomande, T. Ghariani, B. Jouaber, Energy-efficient coordinated beamforming for multi-cell multicast networks under statistical CSI, in: Proceedings of the 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), IEEE, 2017, pp. 1–5.
[24] M. Bengtsson, B. Ottersten, Optimal and Suboptimal Transmit Beamforming, Electrical Engineering & Applied Signal Processing Series, CRC Press, 2001.
[25] T. Van Chien, C. Mollén, E. Björnson, Large-scale-fading decoding in cellular massive mimo systems with spatially correlated channels, IEEE Transactions on Communications (2018) pages 1-1, ISSN: 0090-6778, doi:10.1109/TCOMM.2018.2889090.
[26] S.S. Ioushua, Y.C. Eldar, Pilot contamination mitigation with reduced RF chains, Proceedings of the IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2017, doi:10.1109/spawc.2017.8227727.
[27] J. Zhang, F. Roemer, M. Haardt, Relay assisted physical resource sharing: projection based separation of multiple operators (probasemo) for two-way relaying with MIMO amplify and forward relays, IEEE Trans. Signal Process. 60 (9) (2012) 4834–4848.
[28] M.M. Naghsh, M. Soltanalian, P. Stoica, M. Masjedi, B. Ottersten, Efficient sum-rate maximization for medium-scale MIMO AF-relay networks, IEEE Trans. Wirel. Commun. 15 (9) (2016) 6400–6411.
[29] A. Khabbazibasmenj, F. Roemer, S. Vorobyov, M. Haardt, Sum-rate maximization in two-way AF MIMO relaying: polynomial time solutions to a class of DC programming problems, IEEE Trans. Signal Process. 60 (10) (2012) 5478–5493.
[30] A. Aubry, A. De Maio, A. Farina, M. Wicks, Knowledge-aided (potentially cognitive) transmit signal and receive filter design in signal-dependent clutter, IEEE Trans. Aerosp. Electron. Syst. 49 (2013) 93–117.
[31] F. Gini, A. De Maio, L. Patton, Waveform design and diversity for advanced radar systems, 22, Lond. Inst. Eng. Technol., 2012.
[32] S. Balakrishnama, A. Ganapathiraju, Linear discriminant analysis-a brief tutorial, Inst. Signal inf. Process. 18 (1998).
[33] Y. Huang, Y. Guan, On the linear discriminant analysis for large number of classes, Eng. Appl. Artif. Intell. 43 (2015) 15–26.
[34] H. Cui, R. Li, W. Zhong, Model-free feature screening for ultra-high dimensional discriminant analysis, J. Am. Stat. Assoc 110 (510) (2015) 630–641.
[35] J. Kalina, J.D. Tebbens, Algorithms for regularized linear discriminant analysis, in: Proceedings of the 6th International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS'15), Scitepress, 2015, pp. 128–133.
[36] M.S. Treder, A.K. Porbadnigk, F.S. Avarvand, K.-R. Müller, B. Blankertz, The LDA beamformer: optimal estimation of ERP source time series using linear discriminant analysis, Neuroimage (2016).
[37] H. Mandelkow, J.A. de Zwart, J.H. Duyn, Linear discriminant analysis achieves high classification accuracy for the BOLD fmri response to naturalistic movie stimuli, Front Hum Neurosci. 10 (2016).
[38] A. Sharma, K.K. Paliwal, Linear discriminant analysis for the small sample size problem: an overview, Int. J. Mach. Learn. Cybern. 6 (3) (2015) 443–454.
[39] C.R. Rao, The utilization of multiple measurements in problems of biological classification, J. R. Stat. Soc. Ser B (Methodol.) 10 (2) (1948) 159–203.
[40] S.-J. Kim, A. Magnani, S. Boyd, Robust fisher discriminant analysis, Adv. Neural Inf. Process Syst. 18 (2006) 659.
[41] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, S. Zhang, Semidefinite relaxation of quadratic optimization problems, IEEE Signal Process. Mag. 27 (3) (2010) 20–34.
[42] R.A. Horn, C.R. Johnson, Matrix Analysis, 2 ed., Cambridge University Press, Cambridge; New York, 2012.
[43] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge University Press, 2009.
[44] E. Schechter, Handbook of Analysis and its Foundations, Academic Press, 1996.
[45] P.A. Absil, R. Mahony, R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, 2009.
[46] A. Edelman, T.A. Arias, S.T. Smith, The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl. 20 (2) (1998) 303–353.
[47] E. Stiefel, Richtungsfelder und fernparallelismus in n-dimensionalen mannigfaltigkeiten, Commentarii Mathematici Helvetici 8 (1) (1935) 305–353.
[48] S. Zhang, Y. Huang, Complex quadratic optimization and semidefinite programming, SIAM J. Optim. 16 (3) (2006) 871–890.
[49] M. Soltanalian, P. Stoica, Designing unimodular codes via quadratic optimization, IEEE Trans. Signal Process. 62 (5) (2014) 1221–1234.
[50] M. Soltanalian, B. Tang, J. Li, P. Stoica, Joint design of the receive filter and transmit sequence for active sensing, IEEE Signal Process Lett. 20 (5) (2013) 423–426.
[51] M. Soltanalian, H. Hu, P. Stoica, Single-stage transmit beamforming design for MIMO radar, Signal Process. 102 (2014) 132–138.
[52] S.S. Ioushua, Y.C. Eldar, Hybrid analog-digital beamforming for massive mimo systems, 2017. arXiv preprint, arXiv:1712.03485.
[53] H. Al-Salihi, T.V. Chien, T.A. Le, M.R. Nakhai, A successive optimization approach to pilot design for multi-cell massive mimo systems, IEEE Commun. Lett. 22 (5) (2018) 1086–1089.
[54] W. Yu, T. Lan, Transmitter optimization for the multi-antenna downlink with per-antenna power constraints, IEEE Trans. Signal Process. 55 (6) (2007) 2646–2660.
[55] H. He, J. Li, P. Stoica, Waveform Design for Active sensing Systems: A Computational Approach, Cambridge University Press, 2012.
[56] N. Levanon, E. Mozeson, Radar Signals, Wiley, New York, 2004.
[57] P. Stoica, Y. Selén, Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher, IEEE Signal Process Mag. 21 (1) (2004) 112–114.
[58] D.R. Hunter, K. Lange, A tutorial on MM algorithms, Am. Stat. 58 (1) (2004) 30–37.
[59] E. Feron, Nonconvex quadratic programming, semidefinite relaxations and randomization algorithms in information and decision systems, in: System Theory, Springer, 2000, pp. 255–274.