Designing Unimodular Codes Via Quadratic Optimization

Mojtaba Soltanalian, *Student Member, IEEE*, and Petre Stoica, *Fellow, IEEE*

***Abstract*—The NP-hard problem of optimizing a quadratic form over the unimodular vector set arises in radar code design scenarios as well as other active sensing and communication applications. To tackle this problem (which we call unimodular quadratic program (UQP)), several computational approaches are devised and studied. Power method-like iterations are introduced for local optimization of UQP. Furthermore, a monotonically error-bound improving technique (MERIT) is proposed to obtain the global optimum or a local optimum of UQP with good sub-optimality guarantees. The provided sub-optimality guarantees are case-dependent and may outperform the $\pi/4$ approximation guarantee of semi-definite relaxation. Several numerical examples are presented to illustrate the performance of the proposed method. The examples show that for several cases, including rank-deficient matrices, the proposed methods can solve UQPs efficiently in the sense of sub-optimality guarantee and computational time.**

***Index Terms*—Code design, radar codes, unimodular codes, quadratic programming, peak-to-average-power ratio (PAR).**

I. Introduction

UNIMODULAR codes are used in many active sensing and communication systems mainly as a result of the their optimal (i.e., unity) peak-to-average-power ratio (PAR). The design of such codes can be often formulated as the optimization of a quadratic form (see Sub-section I-A for examples). Therefore, we will study the problem

$$\text{UQP}: \quad \max_{\boldsymbol{s}\in\Omega^n} \boldsymbol{s}^H\boldsymbol{R}\boldsymbol{s} \tag{1}$$

where $\boldsymbol{R}\in\mathbb{C}^{n\times n}$ is a given Hermitian matrix, Ω represents the unit circle, i.e., $\Omega=\{s\in\mathbb{C}: |s|=1\}$ and UQP stands for Unimodular Quadratic Program(ming).

Notation: We use bold lowercase letters for vectors/sequences and bold uppercase letters for matrices. $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ denote the vector/matrix transpose, the complex conjugate, and the Hermitian transpose, respectively. $\mathbf{1}$ and $\mathbf{0}$ are the all-one and all-zero vectors/matrices. $\boldsymbol{e}_k$ is the kth standard basis vector in $\mathbb{C}^n$. $\|\boldsymbol{x}\|_n$ or the l_n-norm of the vector $\boldsymbol{x}$ is defined as $(\sum_k |\boldsymbol{x}(k)|^n)^{\frac{1}{n}}$ where $\{\boldsymbol{x}(k)\}$ are the entries of $\boldsymbol{x}$. The Frobenius norm of a matrix $\boldsymbol{X}$ (denoted by $\|\boldsymbol{X}\|_F$) with entries $\{\boldsymbol{X}(k,l)\}$ is equal to $(\sum_{k,l}|\boldsymbol{X}(k,l)|^2)^{\frac{1}{2}}$. We use $\Re(\boldsymbol{X})$ and $\Im(\boldsymbol{X})$ to denote the matrices obtained by collecting the real parts, and respectively, the imaginary parts of the entries of $\boldsymbol{X}$. The matrix $e^{j\boldsymbol{X}}$ is defined element-wisely as $[e^{j\boldsymbol{X}}]_{k,l}=e^{j[\boldsymbol{X}]_{k,l}}$. $\arg(\cdot)$ denotes the phase angle (in radians) of the vector/matrix argument. $\mathbb{E}[\cdot]$ stands for the expectation operator. $\mathbf{Diag}(\cdot)$ denotes the diagonal matrix formed by the entries of the vector argument, whereas $\mathbf{diag}(\cdot)$ denotes the vector formed by collecting the diagonal entries of the matrix argument. $\sigma_k(\boldsymbol{X})$ represents the kth maximal eigenvalue of $\boldsymbol{X}$. The symbol $\odot$ stands for the Hadamard (element-wise) product of matrices. The operator notation $\oplus$ stands for the Minkowski sum of the two sets. Finally, $\mathbb{R}$ and $\mathbb{C}$ represent the set of real and complex numbers, respectively.

A. Motivating Applications

To motivate the UQP formulation considered above, we present four scenarios in which a design problem in active sensing or communication boils down to an UQP.

• *Designing codes that optimize the SNR or the CRLB:* We consider a monostatic radar which transmits a linearly encoded burst of pulses. The observed backscattered signal $\boldsymbol{v}$ can be written as (see, e.g., [1]):

$$\boldsymbol{v}=a(\boldsymbol{c}\odot\boldsymbol{p})+\boldsymbol{w}, \tag{2}$$

where a represents channel propagation and backscattering effects, $\boldsymbol{w}$ is the disturbance/noise component, $\boldsymbol{c}$ is the unimodular vector containing the code elements, $\boldsymbol{p}=(1,e^{j2\pi f_d T_r},\ldots,e^{j2\pi(n-1)f_d T_r})^T$ is the temporal steering vector with f_d and T_r being the target Doppler frequency and pulse repetition time, respectively.

Under the assumption that $\boldsymbol{w}$ is a zero-mean complex-valued circular Gaussian vector with known positive definite covariance matrix $\mathbb{E}[\boldsymbol{w}\boldsymbol{w}^H]=\boldsymbol{M}$, the signal-to-noise ratio (SNR) is given by [2]

$$\text{SNR}=|a|^2\boldsymbol{c}^H\boldsymbol{R}\boldsymbol{c} \tag{3}$$

where $\boldsymbol{R}=\boldsymbol{M}^{-1}\odot(\boldsymbol{p}\boldsymbol{p}^H)^*$. Therefore, the problem of designing codes optimizing the SNR of the radar system can be formulated directly as an UQP. Additionally, the Cramer-Rao lower bound (CRLB) for the target Doppler frequency estimation (which yields a lower bound on the variance of any unbiased target Doppler frequency estimator) is given by [2]

$$\begin{aligned}\text{CRLB}&=(2|a|^2(\boldsymbol{c}\odot\boldsymbol{p}\odot\boldsymbol{u})^H\boldsymbol{M}^{-1}(\boldsymbol{c}\odot\boldsymbol{p}\odot\boldsymbol{u}))^{-1}\\&=(2|a|^2\boldsymbol{c}^H\boldsymbol{R}'\boldsymbol{c})^{-1}\end{aligned} \tag{4}$$

Manuscript received June 27, 2013; revised October 21, 2013; accepted December 19, 2013. Date of publication January 02, 2014; date of current version February 10, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Joakim Jalden. This work was supported in part by the European Research Council (ERC) under Grant #228044 and the Swedish Research Council.

The authors are with the Department of Information Technology, Uppsala University, Uppsala, SE 75105, Sweden (e-mail: mojtaba.soltanalian@it.uu.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2013.2296883

where $\boldsymbol{u} = (0, j2\pi T_r, \ldots, j2\pi(n-1)T_r)^T$ and $\boldsymbol{R}' = \boldsymbol{M}^{-1} \odot (\boldsymbol{p}\boldsymbol{p}^H)^* \odot (\boldsymbol{u}\boldsymbol{u}^H)^*$. Therefore the minimization of CRLB can also be formulated as an UQP. For the simultaneous optimization of SNR and CRLB see [2].

• *Synthesizing cross ambiguity functions (CAFs):* The ambiguity function (which is widely used in active sensing applications [3], [4]) represents the two-dimensional response of the matched filter to a signal with time delay τ and Doppler frequency shift f. The more general concept of cross ambiguity function occurs when the matched filter is replaced by a mismatched filter. The cross ambiguity function (CAF) is defined as

$$\chi(\tau, f) = \int_{-\infty}^{\infty} u(t)v^*(t+\tau)e^{j2\pi ft}dt \tag{5}$$

where $u(t)$ and $v(t)$ are the transmit signal and the receiver filter, respectively (the ambiguity function is obtained from (5) with $v(t) = u(t)$). In several applications $u(t)$ and $v(t)$ are given by:

$$u(t) = \sum_{k=1}^{n} x_k p_k(t), \quad v(t) = \sum_{k=1}^{n} y_k p_k(t) \tag{6}$$

where $\{p_k(t)\}$ are pulse-shaping functions (with the rectangular pulse as a common example), and

$$\boldsymbol{x} = (x_1 \ldots x_n)^T, \quad \boldsymbol{y} = (y_1 \ldots y_n)^T \tag{7}$$

are the code and, respectively, the filter vectors. The design problem of synthesizing a desired CAF has a small number of free variables (i.e., the entries of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$) compared to the large number of constraints arising from two-dimensional matching criteria (to a given $|\chi(\tau, f)|$). Therefore, the problem is generally considered to be difficult and there are not many methods to synthesize a desired (cross) ambiguity function. Below, we describe briefly the cyclic approach of [5] for CAF design.

The problem of matching a desired $|\chi(\tau, f)| = d(\tau, f)$ can be formulated as the minimization of the criterion [5]

$$g(\boldsymbol{x}, \boldsymbol{y}, \phi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(\tau, f) \times \left| d(\tau, f)e^{j\phi(\tau,f)} - \boldsymbol{y}^H \boldsymbol{J}(\tau, f)\boldsymbol{x} \right|^2 d\tau df \tag{8}$$

where $\boldsymbol{J}(\tau, f) \in \mathbb{C}^{n\times n}$ is given, $w(\tau, f)$ is a weighting function that specifies the CAF area which needs to be emphasized and $\phi(\tau, f)$ represent auxiliary phase variables. It is not difficult to see that for fixed $\boldsymbol{x}$ and $\boldsymbol{y}$, the minimizer $\phi(\tau, f)$ is given by $\phi(\tau, f) = \arg\{\boldsymbol{y}^H \boldsymbol{J}(\tau, f)\boldsymbol{x}\}$. For fixed $\phi(\tau, f)$ and $\boldsymbol{x}$, the criterion g can be written as

$$\begin{aligned} g(\boldsymbol{y}) &= \boldsymbol{y}^H \boldsymbol{D}_1 \boldsymbol{y} - \boldsymbol{y}^H \boldsymbol{B}^H \boldsymbol{x} - \boldsymbol{x}^H \boldsymbol{B}\boldsymbol{y} + \text{const}_1 \\ &= \left(\boldsymbol{y} - \boldsymbol{D}_1^{-1}\boldsymbol{B}^H\boldsymbol{x}\right)^H \boldsymbol{D}_1 \left(\boldsymbol{y} - \boldsymbol{D}_1^{-1}\boldsymbol{B}^H\boldsymbol{x}\right) + \text{const}_2 \end{aligned} \tag{9}$$

where $\boldsymbol{B}$ and $\boldsymbol{D}_1$ are given matrices in $\mathbb{C}^{n\times n}$ [5]. Due to practical considerations, the transmit coefficients $\{x_k\}$ must have low PAR values. However, the receiver coefficients $\{y_k\}$ need not be constrained in such a way. Therefore, the minimizer $\boldsymbol{y}$ of $g(\boldsymbol{y})$ is given by $\boldsymbol{y} = \boldsymbol{D}_1^{-1}\boldsymbol{B}^H\boldsymbol{x}$. Similarly, for fixed $\phi(\tau, f)$ and $\boldsymbol{y}$, the criterion g can be written as

$$g(\boldsymbol{x}) = \boldsymbol{x}^H \boldsymbol{D}_2 \boldsymbol{x} - \boldsymbol{x}^H \boldsymbol{B}\boldsymbol{y} - \boldsymbol{y}^H \boldsymbol{B}^H \boldsymbol{x} + \text{const}_3 \tag{10}$$

where $\boldsymbol{D}_2 \in \mathbb{C}^{n\times n}$ is given [5]. If a unimodular code vector $\boldsymbol{x}$ is desired then the optimization of $g(\boldsymbol{x})$ is an UQP as $g(\boldsymbol{x})$ can be written as

$$g(\boldsymbol{x}) = \begin{pmatrix} e^{j\varphi}\boldsymbol{x} \\ e^{j\varphi} \end{pmatrix}^H \begin{pmatrix} \boldsymbol{D}_2 & -\boldsymbol{B}\boldsymbol{y} \\ -(\boldsymbol{B}\boldsymbol{y})^H & 0 \end{pmatrix} \begin{pmatrix} e^{j\varphi}\boldsymbol{x} \\ e^{j\varphi} \end{pmatrix} + \text{const}_3 \tag{11}$$

where $\varphi \in [0, 2\pi)$ is a free phase variable.

• *Steering vector estimation in adaptive beamforming:* Consider a linear array with n antennas. The output of the array at time instant k can be expressed as [6]

$$\boldsymbol{x}_k = s_k \boldsymbol{a} + \boldsymbol{n}_k \tag{12}$$

with $\{s_k\}$ being the signal waveform, $\boldsymbol{a}$ the associated steering vector (with $|[\boldsymbol{a}]_l| = 1$, $1 \le l \le n$), and $\boldsymbol{n}_k$ the vector accounting for all independent interferences.

The true steering vector is usually unknown in practice, and it can therefore be considered as an unimodular vector to be determined [7]. Define the sample covariance matrix of $\{\boldsymbol{x}_k\}$ as $\hat{\boldsymbol{R}} = \frac{1}{T}\sum_{k=1}^{T} \boldsymbol{x}_k \boldsymbol{x}_k^H$ where T is the number of training data samples. Assuming some prior knowledge on $\boldsymbol{a}$ (which can be represented by $\arg(\boldsymbol{a})$ being in a given sector Θ), the problem of estimating the steering vector can be formulated as [8]

$$\begin{aligned} \min_{\boldsymbol{a}} \quad & \boldsymbol{a}^H \hat{\boldsymbol{R}}^{-1} \boldsymbol{a} \\ \text{s.t.} \quad & \arg(\boldsymbol{a}) \in \Theta, \end{aligned} \tag{13}$$

hence an UQP-type problem. Such problems can be tackled using general local optimization techniques or the optimization scheme introduced in Section III.

• *Maximum likelihood (ML) detection of unimodular codes:* Assume the linear model

$$\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{s} + \boldsymbol{n} \tag{14}$$

where $\boldsymbol{Q}$ represents a multiple-input multiple-output (MIMO) channel, $\boldsymbol{y}$ is the received signal, $\boldsymbol{n}$ is the additive white Gaussian noise and $\boldsymbol{s}$ contains the unimodular symbols which are to be estimated. The ML detection of $\boldsymbol{s}$ may be stated as

$$\hat{\boldsymbol{s}}_{\text{ML}} = \arg\min_{\boldsymbol{s}\in\Omega^n} \|\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{s}\|_2 \tag{15}$$

It is straightforward to verify that the above optimization problem is equivalent to the UQP [10]:

$$\min_{\bar{\boldsymbol{s}}\in\Omega^{n+1}} \bar{\boldsymbol{s}}^H \boldsymbol{R} \bar{\boldsymbol{s}} \tag{16}$$

where

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{Q}^H\boldsymbol{Q} & -\boldsymbol{Q}^H\boldsymbol{y} \\ -\boldsymbol{y}^H\boldsymbol{Q} & 0 \end{pmatrix}, \quad \bar{\boldsymbol{s}} = \begin{pmatrix} e^{j\varphi}\boldsymbol{s} \\ e^{j\varphi} \end{pmatrix} \tag{17}$$

and where $\varphi \in [0, 2\pi)$ is a free phase variable.

B. Related Work

In [11], the NP-hardness of UQP is proven by employing a reduction from an NP-complete matrix partitioning problem. The UQP in (1) is often studied along with the following (also NP-hard) related problem in which the decision variables are discrete:

$$m\text{-UQP}: \max_{\boldsymbol{s}\in\Omega_m^n} \boldsymbol{s}^H\boldsymbol{R}\boldsymbol{s} \tag{18}$$

where $\Omega_m = \{1, e^{j\frac{2\pi}{m}}, \ldots, e^{j\frac{2\pi}{m}(m-1)}\}$. Note that the latter problem coincides with the UQP in (1) as $m \to \infty$. The authors of [12] show that when the matrix $\boldsymbol{R}$ is rank-deficient (more precisely, when $d = \mathrm{rank}(\boldsymbol{R})$ behaves like $\mathcal{O}(1)$ with respect to the problem dimension) the m-UQP problem can be solved in polynomial-time and they propose a $\mathcal{O}((mn/2)^{2d})$-complexity algorithm to solve (18). However, such algorithms are not applicable to the UQP which corresponds to $m \to \infty$.

Studies on polynomial-time (or efficient) algorithms for UQP (and m-UQP) have been extensive (e.g., see [9]–[22] and the references therein). In particular, the semi-definite relaxation (SDR) related techniques have been the most appealing approaches to the researchers. To derive an SDR, we note that $\boldsymbol{s}^H\boldsymbol{R}\boldsymbol{s} = \mathrm{tr}(\boldsymbol{s}^H\boldsymbol{R}\boldsymbol{s}) = \mathrm{tr}(\boldsymbol{R}\boldsymbol{s}\boldsymbol{s}^H)$. Hence, the UQP can be rewritten as

$$\begin{aligned} &\max_{\boldsymbol{S}} \mathrm{tr}(\boldsymbol{R}\boldsymbol{S}) \\ &\text{s.t. } \boldsymbol{S} = \boldsymbol{s}\boldsymbol{s}^H, \boldsymbol{s} \in \Omega^n. \end{aligned} \tag{19}$$

If we relax (19) by removing the rank constraint on $\boldsymbol{S}$ then the result is a semi-definite program:

$$\begin{aligned} \mathrm{SDP}: \quad &\max_{\boldsymbol{S}} \mathrm{tr}(\boldsymbol{R}\boldsymbol{S}) \\ &\text{s.t. } [\boldsymbol{S}]_{k,k} = 1, \quad 1 \le k \le n, \\ &\boldsymbol{S} \text{ is positive semi-definite.} \end{aligned} \tag{20}$$

The above SDP can be solved in polynomial time using interior-point methods [17]. The approximation of the UQP solution based on the SDP solution can be accomplished in several ways. For example, we can approximate the phase values of the solution $\boldsymbol{s}$ using a rank-one approximation of $\boldsymbol{S}$. A more effective approach for guessing $\boldsymbol{s}$ is based on randomized approximations (see [11], [18] and [19]). A detailed guideline for randomized approximation of the UQP solution can be found in [19]. In addition, we refer the interested reader to the survey of the rich literature on SDR in [20].

In order to formalize the quality assessment of the UQP solutions, let $\boldsymbol{s}$ be the approximate solution to a given UQP. We assume that $\boldsymbol{R}$ is positive semidefinite (such an assumption can be made without loss of generality, see Section II-A). Then the *approximation ratio* (δ) associated with $\boldsymbol{s}$ is given by

$$\delta \triangleq \frac{\boldsymbol{s}^H\boldsymbol{R}\boldsymbol{s}}{\max_{\boldsymbol{s}'\in\Omega^n} \boldsymbol{s}'^H\boldsymbol{R}\boldsymbol{s}'}. \tag{21}$$

The approximation ratio is usually unknown, because the global optimum of the problem is not known. However, an optimization method may offer a *sub-optimality guarantee* (γ), i.e., a lower bound on the quality of the approximate solution:

$$\delta \ge \gamma. \tag{22}$$

Herein, we present the existing (analytically derived) sub-optimality guarantee for SDR. Let v_{SDR} be the expected value of the UQP objective at the obtained randomized solution. Let v_{opt} represent the optimal value of the UQP objective. We have

$$\gamma v_{\mathrm{opt}} \le v_{\mathrm{SDR}} \le v_{\mathrm{opt}} \tag{23}$$

with the sub-optimality guarantee coefficient $\gamma = \pi/4$ [11], [21]. Note that the sub-optimality coefficient of the solution obtained by SDR can be arbitrarily close to $\pi/4$ (e.g., see [21]). For the sake of brevity, in the sequel the abbreviation SDR will be used for semidefinite relaxations followed by the randomization procedure.

C. Contributions of This Work

Besides SDR, the literature does not offer many other numerical frameworks to tackle UQP. In this paper, a specialized local optimization scheme for UQP is proposed. The proposed computationally efficient local optimization approach can be used to tackle UQP as well as improve upon the solutions obtained by other methods such as SDR. Furthermore, a **m**onotonically **er**ror-bound **i**mproving **t**echnique (called MERIT) is introduced to obtain the global optimum or a local optimum of UQP with good sub-optimality guarantees. Note that:

- MERIT provides real-time case-dependent sub-optimality guarantees (γ) during its iterations. To the best of our knowledge, such guarantees for UQP were not known prior to this work. Using MERIT one may obtain better performance guarantees compared to the analytical worst-case guarantees (such as $\gamma = \pi/4$ for SDR).
- The provided case-dependent sub-optimality guarantees are of practical importance in decision making scenarios. For instance in some cases the UQP solution obtained by SDR (or other optimization methods) might achieve good objective values, and equivalently good approximation ratios δ (this is indeed the case for some practical examples, see Section VI). However, unless the goodness of the obtained solution is known (which can be determined using the proposed bounds), the solution cannot be trusted.
- Using MERIT, numerical evidence is provided to show that several UQPs (particularly those with low rank) can be solved efficiently without sacrificing the solution accuracy.

Finally, we believe that the general ideas of this work can be adopted to tackle m-UQP as the finite alphabet case of UQP. However, a detailed study of m-UQP is beyond the scope of this paper.

The rest of this work is organized as follows. Section II discusses several properties of UQP. Section III introduces a specialized local optimization method that resembles the well-known power method. Section IV presents a cone approximation that is used in Section V to derive the algorithmic form of MERIT for UQP. Several numerical examples are provided in Section VI. Finally, Section VII concludes the paper.

II. Some Properties of UQP

In this section, we study several properties of UQP. The discussed properties lay the grounds for a better understanding of UQP as well as the tools proposed to tackle it in the following sections.

A. Basic Properties

The UQP formulation in (1) covers both maximization and minimization of quadratic forms (one can obtain the minimization of the quadratic form in (1) by considering $-\boldsymbol{R}$ in lieu of $\boldsymbol{R}$). In addition, without loss of generality, the Hermitian matrix $\boldsymbol{R}$ can be assumed to be positive (semi)definite. If $\boldsymbol{R}$ is not positive (semi)definite, we can make it so using the diagonal loading technique (i.e., $\boldsymbol{R} \leftarrow \boldsymbol{R} + \lambda \boldsymbol{I}$ where $\lambda \geq -\sigma_n(\boldsymbol{R})$). Note that such a diagonal loading does not change the solution of UQP as $\boldsymbol{s}^H(\boldsymbol{R} + \lambda \boldsymbol{I})\boldsymbol{s} = \boldsymbol{s}^H \boldsymbol{R} \boldsymbol{s} + \lambda n$. Next, we note that if $\tilde{\boldsymbol{s}}$ is a solution to UQP then $e^{j\phi}\tilde{\boldsymbol{s}}$ (for any $\phi \in [0, 2\pi)$) is also a valid solution. To establish connections among different UQPs, Theorem 1 presents a bijection among the set of matrices leading to the same solution.

Theorem 1: Let $\mathcal{K}(\boldsymbol{s})$ represent the set of matrices $\boldsymbol{R}$ for which a given $\boldsymbol{s} \in \Omega^n$ is the global optimizer of UQP. Then

1) $\mathcal{K}(\boldsymbol{s})$ is a convex cone.
2) For any two vectors $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \Omega^n$, the one-to-one mapping (where $\boldsymbol{s}_0 = \boldsymbol{s}_1^* \odot \boldsymbol{s}_2$)

$$\boldsymbol{R} \in \mathcal{K}(\boldsymbol{s}_1) \Longleftrightarrow \boldsymbol{R} \odot \left(\boldsymbol{s}_0 \boldsymbol{s}_0^H\right) \in \mathcal{K}(\boldsymbol{s}_2) \tag{24}$$

holds among the matrices in $\mathcal{K}(\boldsymbol{s}_1)$ and $\mathcal{K}(\boldsymbol{s}_2)$.

Proof: See the Appendix. ∎

It is interesting to note that in light of the above result, the characterization of the cone $\mathcal{K}(\boldsymbol{s})$ for any given $\boldsymbol{s} = \tilde{\boldsymbol{s}}$ leads to a complete characterization of all $\mathcal{K}(\boldsymbol{s})$, $\boldsymbol{s} \in \Omega^n$, and thus solving any UQP. However, the NP-hardness of UQP suggests that such a tractable characterization cannot be expected. Further discussions regarding the characterization of $\mathcal{K}(\boldsymbol{s})$ are deferred to Section IV.

B. Analytical Solutions to UQP

There exist cases for which the analytical global optima of UQP are easy to obtain. In this sub-section, we consider two such cases which will be used in Section IV to provide an approximate characterization of $\mathcal{K}(\boldsymbol{s})$. A special example is the case in which $e^{j\arg(\boldsymbol{R})}$ (see the notation definition in the Introduction) is a rank-one matrix. More precisely, let $\boldsymbol{R} = \boldsymbol{R}_1 \odot (\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}^H)$ where $\boldsymbol{R}_1$ is a real-valued Hermitian matrix with nonnegative entries and $\tilde{\boldsymbol{s}} \in \Omega^n$. A simple special case of this example is when $\boldsymbol{R}$ is a rank-one matrix itself. In this case, it can be easily verified that $\boldsymbol{R}_1 \in \mathcal{K}(\boldsymbol{1}_{n\times 1})$. Therefore, using Theorem 1 one concludes that $\boldsymbol{R} \in \mathcal{K}(\tilde{\boldsymbol{s}})$ i.e., $\boldsymbol{s} = \tilde{\boldsymbol{s}}$ yields the global optimum of UQP. As another example, Theorem 2 considers the case for which the matrix $\boldsymbol{R}$ has a repeated largest eigenvalue.

Theorem 2: Let $\boldsymbol{R}$ be a Hermitian matrix with eigenvalue decomposition $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H$. Suppose $\boldsymbol{\Sigma}$ is of the form

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathbf{Diag}([\underbrace{\sigma_1 \cdots \sigma_1}_{m \text{ times}} \sigma_2 \cdots \sigma_{n-m+1}]^T) \\ \sigma_1 &> \sigma_2 \geq \cdots \geq \sigma_{n-m+1} \end{aligned} \tag{25}$$

and let $\boldsymbol{U}_m$ be the matrix made from the first m columns of $\boldsymbol{U}$. Now suppose $\tilde{\boldsymbol{s}} \in \Omega^n$ lies in the linear space spanned by the columns of $\boldsymbol{U}_m$, i.e., there exists a vector $\boldsymbol{\alpha} \in \mathbb{C}^m$ such that

$$\tilde{\boldsymbol{s}} = \boldsymbol{U}_m \boldsymbol{\alpha}. \tag{26}$$

Then $\tilde{\boldsymbol{s}}$ is a global optimizer of UQP.

Proof: If $\tilde{\boldsymbol{s}}$ satisfies (24), then it belongs to the span of the m dominant eigenvectors of $\boldsymbol{R}$, and hence it is also a dominant eigenvector of $\boldsymbol{R}$. This fact implies that $\tilde{\boldsymbol{s}}$ is the global optimizer of the quadratic optimization (even without the unimodularity constraint) which completes the proof. ∎

We end this section by noting that the solution to an UQP is not necessarily unique. For any set of unimodular vectors $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k\}$, $k \leq n$, we can use the Gram-Schmidt process to obtain a unitary matrix $\boldsymbol{U}$ the first k columns of which span the same linear space as $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k$. In this case, Theorem 2 suggests a method to construct a matrix $\boldsymbol{R}$ (by choosing a $\boldsymbol{\Sigma}$ with k identical largest eigenvalues) for which all $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k$ are global optimizers of the corresponding UQP.

III. Power Method for UQP

Due to its NP-hard nature, UQP has in general a highly multimodal optimization objective. Finding the local optima of UQP is not only useful to tackle the problem itself (particularly for UQP-related problems such as (13)), but also to improve the UQP approximate solutions obtained by SDR or other optimization techniques. In this section, we introduce a computationally efficient procedure (to obtain a local optimum of UQP) which resembles the well-known power method for computing the dominant eigenvalue/vector pairs of matrices.

Assume that $\boldsymbol{R}$ is positive definite and let $\{\boldsymbol{s}^{(t+1)}\}_{t=0}^{\infty}$ be a sequence of unimodular codes where $\boldsymbol{s}^{(t+1)}$ is the minimizer of the following criterion:

$$\min_{\boldsymbol{s}^{(t+1)} \in \Omega^n} \left\| \boldsymbol{s}^{(t+1)} - \boldsymbol{R}\boldsymbol{s}^{(t)} \right\|_2 \tag{27}$$

The minimizing vector $\boldsymbol{s}^{(t+1)}$ of (27) is simply given by the following *power method-like* iteration:

$$\boldsymbol{s}^{(t+1)} = e^{j\arg(\boldsymbol{R}\boldsymbol{s}^{(t)})} \tag{28}$$

Note that

$$\left\| \boldsymbol{s}^{(t+1)} - \boldsymbol{R}\boldsymbol{s}^{(t)} \right\|_2^2 = \text{const} - 2\Re\left\{ \boldsymbol{s}^{(t+1)\,H} \boldsymbol{R}\, \boldsymbol{s}^{(t)} \right\} \tag{29}$$

As a result, $\boldsymbol{s}^{(t+1)}$ is equivalently the maximizer of the criterion $\Re\{\boldsymbol{s}^{(t+1)\,H} \boldsymbol{R}\, \boldsymbol{s}^{(t)}\}$. Moreover, if $\boldsymbol{s}^{(t+1)} \neq \boldsymbol{s}^{(t)}$ we have that

$$\left(\boldsymbol{s}^{(t+1)} - \boldsymbol{s}^{(t)}\right)^H \boldsymbol{R} \left(\boldsymbol{s}^{(t+1)} - \boldsymbol{s}^{(t)}\right) > 0 \tag{30}$$

which implies

$$\begin{aligned} \boldsymbol{s}^{(t+1)\,H} \boldsymbol{R}\, \boldsymbol{s}^{(t+1)} &> 2\Re\left\{ \boldsymbol{s}^{(t+1)\,H} \boldsymbol{R}\, \boldsymbol{s}^{(t)} \right\} - \boldsymbol{s}^{(t)\,H} \boldsymbol{R}\, \boldsymbol{s}^{(t)} \\ &> \boldsymbol{s}^{(t)\,H} \boldsymbol{R}\, \boldsymbol{s}^{(t)} \end{aligned} \tag{31}$$

as $\Re\{\boldsymbol{s}^{(t+1)\,H} \boldsymbol{R}\, \boldsymbol{s}^{(t)}\} > \boldsymbol{s}^{(t)\,H} \boldsymbol{R}\, \boldsymbol{s}^{(t)}$. Therefore, the UQP objective is increasing through the power method-like iterations in (28). On the other hand, the UQP objective is upper bounded by $\sum_{k,l} |\boldsymbol{R}(k,l)|$, and thus the said iterations are convergent in the sense of the UQP objective value. We further note that the

increase in the UQP objective is lower bounded (within a multiplicative constant) by the l_2-norm of the difference of the unimodular codes in successive iterations, viz.

$$\begin{aligned}
&\boldsymbol{s}^{(t+1)\,H}\boldsymbol{R}\,\boldsymbol{s}^{(t+1)} - \boldsymbol{s}^{(t)\,H}\boldsymbol{R}\,\boldsymbol{s}^{(t)} \\
&\quad = \left(\boldsymbol{s}^{(t+1)} - \boldsymbol{s}^{(t)}\right)^H \boldsymbol{R}\left(\boldsymbol{s}^{(t+1)} - \boldsymbol{s}^{(t)}\right) \\
&\qquad + 2\,\Re\left\{\boldsymbol{s}^{(t+1)\,H}\boldsymbol{R}\,\boldsymbol{s}^{(t)}\right\} - 2\,\boldsymbol{s}^{(t)\,H}\boldsymbol{R}\,\boldsymbol{s}^{(t)} \\
&\quad > \left(\boldsymbol{s}^{(t+1)} - \boldsymbol{s}^{(t)}\right)^H \boldsymbol{R}\left(\boldsymbol{s}^{(t+1)} - \boldsymbol{s}^{(t)}\right) \\
&\quad \geq \sigma_n(\boldsymbol{R})\left\|\boldsymbol{s}^{(t+1)} - \boldsymbol{s}^{(t)}\right\|_2^2
\end{aligned} \tag{32}$$

Due to the fact that the sequence $\{\boldsymbol{s}^{(t)\,H}\boldsymbol{R}\,\boldsymbol{s}^{(t)}\}$ is convergent, (32) implies that $\|\boldsymbol{s}^{(t+1)} - \boldsymbol{s}^{(t)}\|_2$ is also converging to zero through the iterations in (28).

It is also important to observe that the power method-like iterations do not stop before reaching a local optimum or saddle point of UQP. A limit point $\tilde{\boldsymbol{s}}$ of (28) can be characterized by the equation

$$\boldsymbol{R}\tilde{\boldsymbol{s}} = \boldsymbol{v} \odot \tilde{\boldsymbol{s}} \tag{33}$$

where $\boldsymbol{v}$ is real-valued and non-negative. On the other hand, the stationary points $\tilde{\boldsymbol{s}}$ of UQP (associated with $\boldsymbol{R}$) may be characterized as $\boldsymbol{R}\tilde{\boldsymbol{s}} = \boldsymbol{v} \odot \tilde{\boldsymbol{s}}$, where $\boldsymbol{v}$ is real-valued (see Appendix B for a detailed derivation). Therefore, the limit points of (28) form a subset of the stationary points of UQP. We refer to the subset of UQP stationary points satisfying (33) as the *stable points* of UQP. A characterization of the UQP optima can also be found in Appendix B. Namely, $\boldsymbol{s}$ is a local maximum of UQP if and only if $\boldsymbol{V} \geq \boldsymbol{R}$, where $\boldsymbol{V} = \mathbf{Diag}(\boldsymbol{v})$. Due to the positive definiteness of $\boldsymbol{R}$, the latter condition implies that for any local maximum of UQP $\boldsymbol{v}$ is non-negative. As a result, the set of the local maxima of UQP (including its global optima) is simply a subset of the stable points of UQP.

Remark 1: The application of the power method-like iterations introduced above is not limited to the optimization of quadratic forms over the unimodular vector set. If one can minimize the criterion in (27) for a particular constraint on $\boldsymbol{s}^{(t+1)}$, say $\boldsymbol{s}^{(t+1)} \in \Psi$, then all the arguments accompanying (27)–(32) are valid and they yield an optimization of quadratic forms over Ψ. An interesting practical example is the more general problem of quadratic optimization over PAR constrained codes (see e.g., [19] and [23]) that can be cast as

$$\begin{aligned}
\max_{\boldsymbol{s}} \quad & \boldsymbol{s}^H\boldsymbol{R}\boldsymbol{s} \\
\text{s.t.} \quad & |\boldsymbol{s}(k)| \leq \sqrt{\gamma},\ \forall\, k, \\
& \|\boldsymbol{s}\|_2^2 = n.
\end{aligned} \tag{34}$$

where γ denotes the maximal tolerable PAR value. The related power method-like iterations, namely

$$\begin{aligned}
\min_{\boldsymbol{s}^{(t+1)}} \quad & \left\|\boldsymbol{s}^{(t+1)} - \boldsymbol{R}\,\boldsymbol{s}^{(t)}\right\|_2 \\
\text{s.t.} \quad & \left|\boldsymbol{s}^{(t+1)}(k)\right| \leq \sqrt{\gamma},\ \forall\, k, \\
& \left\|\boldsymbol{s}^{(t+1)}\right\|_2^2 = n.
\end{aligned} \tag{35}$$

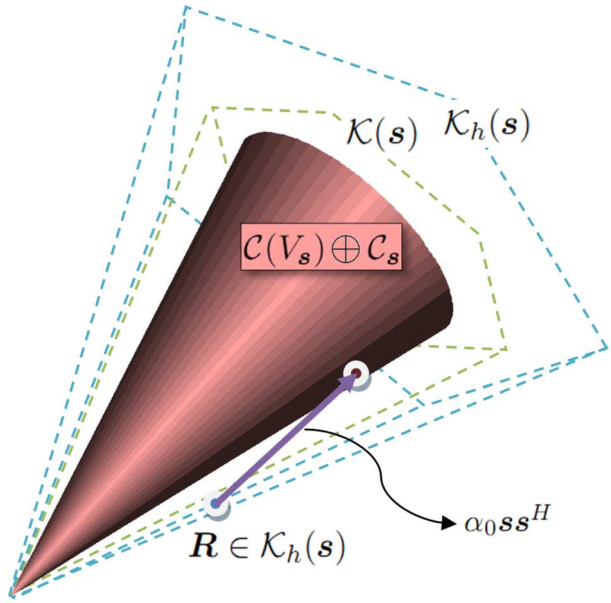


Fig. 1. An illustration of the result in Theorem 3. $\mathcal{K}_h(\boldsymbol{s})$ denotes the convex cone of matrices with $\boldsymbol{s}$ as a stable point of the associated UQPs.

are *nearest-vector* problems that can be solved efficiently via an algorithm devised in [24]. ■

IV. Results on the Cone $\mathcal{K}(\boldsymbol{s})$

While a complete tractable characterization of $\mathcal{K}(\boldsymbol{s})$ cannot be expected (due to the NP-hardness of UQP), approximate characterizations of $\mathcal{K}(\boldsymbol{s})$ are possible. The goal of this section is to provide an approximate characterization of the cone $\mathcal{K}(\boldsymbol{s})$ which can be used to tackle the UQP problem.

Theorem 3: For any given $\boldsymbol{s} = (e^{j\phi_1}, \ldots, e^{j\phi_n})^T \in \Omega^n$, let $\mathcal{C}(\boldsymbol{V_s})$ represent the convex cone of matrices $\boldsymbol{V_s} = \boldsymbol{D} \odot (\boldsymbol{s}\boldsymbol{s}^H)$ where $\boldsymbol{D}$ is any real-valued symmetric matrix with non-negative off-diagonal entries. Also let $\mathcal{C}_{\boldsymbol{s}}$ represent the convex cone of matrices with $\boldsymbol{s}$ being their dominant eigenvector (i.e., the eigenvector corresponding to the maximal eigenvalue). Then for any $\boldsymbol{R} \in \mathcal{K}(\boldsymbol{s})$, there exists $\alpha_0 \geq 0$ such that for all $\alpha \geq \alpha_0$,

$$\boldsymbol{R} + \alpha\boldsymbol{s}\boldsymbol{s}^H \in \mathcal{C}(\boldsymbol{V_s}) \oplus \mathcal{C}_{\boldsymbol{s}}. \tag{36}$$

The proof of Theorem 3 will be presented in several steps (Theorems 4–7 and thereafter). As indicated earlier, a global optimum of UQP is also a stable point of UQP. In what follows, we prove Theorem 3 by proving a more general result, namely that (36) is also satisfied if $\boldsymbol{s}$ is a stable point of UQP (characterized by (33)). However, since $\boldsymbol{s}$ is the global optimum of UQP for all matrices in $\mathcal{C}_{\boldsymbol{s}}$ and $\mathcal{C}(\boldsymbol{V_s})$, the case of $\alpha_0 = 0$ can occur only when $\boldsymbol{s}$ is a global optimum of UQP associated with $\boldsymbol{R}$. An intuitive illustration of the result in Theorem 3 is shown in Fig. 1.

Suppose $\boldsymbol{s}$ is a stable point of UQP associated with a given positive definite matrix $\boldsymbol{R}$, and let $\theta_{k,l} = [\arg(\boldsymbol{R})]_{k,l}$. We define the matrix $\boldsymbol{R}_+^{\boldsymbol{s}}$ as

$$\boldsymbol{R}_+^{\boldsymbol{s}}(k,l) = \begin{cases} |\boldsymbol{R}(k,l)|\cos(\theta_{k,l} - (\phi_k - \phi_l)) & (k,l) \in \Theta, \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

where Θ represents the set of all (k,l) such that $|\theta_{k,l}-(\phi_k-\phi_l)|<\pi/2$. Now, let ρ be a positive real number such that

$$\rho > \max_{(k,l)\notin\Theta}\{|\boldsymbol{R}(k,l)\cos(\theta_{k,l}-(\phi_k-\phi_l))|\} \tag{38}$$

and consider the sequence of matrices $\{\boldsymbol{R}^{(t)}\}$ defined (in an iterative manner) by $\boldsymbol{R}^{(0)}=\boldsymbol{R}$, and

$$\boldsymbol{R}^{(t+1)} = \boldsymbol{R}^{(t)} - \left(\boldsymbol{R}_+^{\boldsymbol{s}^{(t)}} - \rho\mathbf{1}_{n\times n}\right)\odot(\boldsymbol{s}\boldsymbol{s}^H) \tag{39}$$

for $t\geq 0$. The next two theorems (whose proofs are given in the Appendix) study some useful properties of the sequence $\{\boldsymbol{R}^{(t)}\}$.

Theorem 4: $\{\boldsymbol{R}^{(t)}\}$ is convergent in at most two iterations:

$$\boldsymbol{R}^{(t)} = \boldsymbol{R}^{(2)}, \quad \forall\, t\geq 2. \tag{40}$$

Theorem 5: $\boldsymbol{R}^{(t)}$ is a function of ρ. Let ρ and ρ' both satisfy the criterion (38). At the convergence of $\{\boldsymbol{R}^{(t)}\}$ (which is attained for $t=2$) we have:

$$\boldsymbol{R}^{(2)}(\rho') = \boldsymbol{R}^{(2)}(\rho) + (\rho'-\rho)(\boldsymbol{s}\boldsymbol{s}^H). \tag{41}$$

Using the above results, Theorems 6 (whose proof is given in the Appendix) and 7 pave the way for a constructive proof of Theorem 3.

Theorem 6: If $\boldsymbol{s}$ is a stable point of the UQP associated with $\boldsymbol{R}^{(0)}=\boldsymbol{R}$ then it is also a stable point of the UQPs associated with $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$. Furthermore, $\boldsymbol{s}$ is an eigenvector of $\boldsymbol{R}^{(2)}$ corresponding to the eigenvalue $n\rho$.

Theorem 7: If $\boldsymbol{s}$ is a stable point of UQP for $\boldsymbol{R}^{(0)}=\boldsymbol{R}$ then it will be the dominant eigenvector of $\boldsymbol{R}^{(2)}$ if ρ is sufficiently large. In particular, let μ be the largest eigenvalue of $\boldsymbol{R}^{(2)}$ which belongs to an eigenvector other than $\boldsymbol{s}$. Then for any $\rho\geq\mu/n$, $\boldsymbol{s}$ is a dominant eigenvector of $\boldsymbol{R}^{(2)}$.

Proof: We know from Theorem 6 that $\boldsymbol{s}$ is an eigenvector of $\boldsymbol{R}^{(2)}$ corresponding to the eigenvalue $n\rho$. However, if $\boldsymbol{s}$ is not the dominant eigenvector of $\boldsymbol{R}^{(2)}$, Theorem 5 implies that increasing ρ would not change any of the eigenvalues/vectors of $\boldsymbol{R}^{(2)}$ except that it increases the eigenvalue corresponding to $\boldsymbol{s}$. As a result, for $\boldsymbol{s}$ to be the dominant eigenvector of $\boldsymbol{R}^{(2)}$ we only need ρ to satisfy $n\rho\geq\mu$ or equivalently $\rho\geq\mu/n$, which concludes the proof. ∎

Returning to Theorem 3, note that $\boldsymbol{R}$ can be written as

$$\begin{aligned}\boldsymbol{R} &= \boldsymbol{R}^{(0)}\\ &= \boldsymbol{R}^{(2)} + \left(\boldsymbol{R}_+^{\boldsymbol{s}^{(0)}} + \boldsymbol{R}_+^{\boldsymbol{s}^{(1)}}\right)\odot(\boldsymbol{s}\boldsymbol{s}^H) - 2\rho\boldsymbol{s}\boldsymbol{s}^H.\end{aligned} \tag{42}$$

For sufficiently large ρ (satisfying both (38) and the condition of Theorem 7) we have that

$$\boldsymbol{R}+2\rho\boldsymbol{s}\boldsymbol{s}^H = \boldsymbol{R}^{(2)} + \left(\boldsymbol{R}_+^{\boldsymbol{s}^{(0)}} + \boldsymbol{R}_+^{\boldsymbol{s}^{(1)}}\right)\odot(\boldsymbol{s}\boldsymbol{s}^H) \tag{43}$$

where $\boldsymbol{R}^{(2)}\in\mathcal{C}_{\boldsymbol{s}}$ and $(\boldsymbol{R}_+^{\boldsymbol{s}^{(0)}}+\boldsymbol{R}_+^{\boldsymbol{s}^{(1)}})\odot(\boldsymbol{s}\boldsymbol{s}^H)\in\mathcal{C}(\boldsymbol{V}_{\boldsymbol{s}})$. Theorem 3 can thus be directly satisfied using (43) with $\alpha_0=2\rho$.

We conclude this section with two remarks. First of all, the above proof of Theorem 3 does not attempt to derive the minimal α_0. In the following section we study a computational method to obtain an α_0 which is as small as possible. Secondly, we can use $\mathcal{C}(\boldsymbol{V}_{\boldsymbol{s}})\oplus\mathcal{C}_{\boldsymbol{s}}$ as an approximate characterization of $\mathcal{K}(\boldsymbol{s})$ noting that the accuracy of such a characterization can be measured by the minimal value of α_0. An explicit formulation of a sub-optimality guarantee for a solution of UQP based on the above $\mathcal{K}(\boldsymbol{s})$ approximation is derived in the following section.

V. Merit for UQP

Using the previous results, namely the one-to-one mapping introduced in Theorem 1 and the approximation of $\mathcal{K}(\boldsymbol{s})$ derived in Section IV, we build a sequence of matrices (for which the UQP global optima are known) whose distance from a given matrix is decreasing. The proposed iterative approach can be used to solve for the global optimum of UQP or at least to obtain a local optimum (with an upper bound on the sub-optimality of the solution). The sub-optimality guarantees are derived noting that the proposed method decreases an upper bound on the sub-optimality of the obtained UQP solution in each iteration.

We know from Theorem 3 that if $\boldsymbol{s}$ is a stable point of the UQP associated with $\boldsymbol{R}$ then there exist matrices $\boldsymbol{Q}_{\boldsymbol{s}}\in\mathcal{C}_{\boldsymbol{s}}$, $\boldsymbol{P}_{\boldsymbol{s}}\in\mathcal{C}(\boldsymbol{V}_{\boldsymbol{s}})$ and a scalar $\alpha_0\geq 0$ such that

$$\boldsymbol{R}+\alpha_0\boldsymbol{s}\boldsymbol{s}^H = \boldsymbol{Q}_{\boldsymbol{s}}+\boldsymbol{P}_{\boldsymbol{s}}. \tag{44}$$

Equation (44) can be rewritten as

$$\boldsymbol{R}+\alpha_0\boldsymbol{s}\boldsymbol{s}^H = (\boldsymbol{Q}_{\mathbf{1}}+\boldsymbol{P}_{\mathbf{1}})\odot(\boldsymbol{s}\boldsymbol{s}^H) \tag{45}$$

where $\boldsymbol{Q}_{\mathbf{1}}\in\mathcal{C}_{\mathbf{1}}$, $\boldsymbol{P}_{\mathbf{1}}\in\mathcal{C}(\boldsymbol{V}_{\mathbf{1}})$. We first consider the case of $\alpha_0=0$ which corresponds to the global optimality of $\boldsymbol{s}$.

A. Global Optimization of UQP (the Case of $\alpha_0=0$)

Consider the optimization problem:

$$\min_{\boldsymbol{s}\in\Omega^n,\boldsymbol{Q}_{\mathbf{1}}\in\mathcal{C}_{\mathbf{1}},\boldsymbol{P}_{\mathbf{1}}\in\mathcal{C}(\boldsymbol{V}_{\mathbf{1}})} \|\boldsymbol{R}-(\boldsymbol{Q}_{\mathbf{1}}+\boldsymbol{P}_{\mathbf{1}})\odot(\boldsymbol{s}\boldsymbol{s}^H)\|_F \tag{46}$$

Note that, as $\mathcal{C}_{\mathbf{1}}\oplus\mathcal{C}(\boldsymbol{V}_{\mathbf{1}})$ is a convex cone, the global optimizers $\boldsymbol{Q}_{\mathbf{1}}$ and $\boldsymbol{P}_{\mathbf{1}}$ of (46) for any given $\boldsymbol{s}$ can be easily found. On the other hand, the problem of finding an optimal $\boldsymbol{s}$ for fixed $\boldsymbol{R}_{\mathbf{1}}=\boldsymbol{Q}_{\mathbf{1}}+\boldsymbol{P}_{\mathbf{1}}$ is non-convex and hence more difficult to solve globally (see below for details).

We will assume that $\boldsymbol{R}_{\mathbf{1}}$ is a positive definite matrix. To justify this assumption let $\bar{\boldsymbol{R}}=\boldsymbol{R}\odot(\boldsymbol{s}\boldsymbol{s}^H)^*$ and note that the eigenvalues of $\bar{\boldsymbol{R}}$ are exactly the same as those of $\boldsymbol{R}$, hence $\bar{\boldsymbol{R}}$ is positive definite. Suppose that we have

$$\begin{cases}\boldsymbol{x}^H\bar{\boldsymbol{R}}\boldsymbol{x}>\varepsilon, \quad \forall\ \text{unit-norm}\ \boldsymbol{x}\in\mathbb{C}^{n\times 1}\\ \|\bar{\boldsymbol{R}}-\boldsymbol{R}_{\mathbf{1}}\|_F\leq\varepsilon\end{cases} \tag{47}$$

for some $\varepsilon\geq 0$. It follows from (47) that

$$\begin{aligned}\boldsymbol{x}^H\boldsymbol{R}_{\mathbf{1}}\boldsymbol{x} &\geq \boldsymbol{x}^H\bar{\boldsymbol{R}}\boldsymbol{x} - |\boldsymbol{x}^H\bar{\boldsymbol{R}}\boldsymbol{x}-\boldsymbol{x}^H\boldsymbol{R}_{\mathbf{1}}\boldsymbol{x}|\\ &> \varepsilon - |\boldsymbol{x}^H(\bar{\boldsymbol{R}}-\boldsymbol{R}_{\mathbf{1}})\boldsymbol{x}|\\ &\geq \varepsilon - |\sigma_1(\bar{\boldsymbol{R}}-\boldsymbol{R}_{\mathbf{1}})|\\ &\geq \varepsilon - \|\bar{\boldsymbol{R}}-\boldsymbol{R}_{\mathbf{1}}\|_F \geq 0\end{aligned} \tag{48}$$

which implies that $\boldsymbol{R}_{\mathbf{1}}$ is also a positive definite matrix. The conditions in (47) can be met as follows. By considering the partial minimization of (46) only with respect to the component of $\boldsymbol{R}_{\mathbf{1}}$ in $\mathcal{C}(\boldsymbol{V}_{\mathbf{1}})$ (namely $\boldsymbol{P}_{\mathbf{1}}$) we observe that any positive (i.e., with $\lambda>0$) diagonal loading of $\boldsymbol{R}$, which leads to the same

diagonal loading of $\bar{\boldsymbol{R}}$ (as $\bar{\boldsymbol{R}} + \lambda \boldsymbol{I} = \boldsymbol{R} \odot (\boldsymbol{s}\boldsymbol{s}^H)^* + \lambda \boldsymbol{I} = (\boldsymbol{R} + \lambda \boldsymbol{I}) \odot (\boldsymbol{s}\boldsymbol{s}^H)^*$), will be absorbed[1] in $\boldsymbol{P_1}$. Therefore, a positive diagonal loading of $\boldsymbol{R}$ does not change $\|\bar{\boldsymbol{R}} - \boldsymbol{R_1}\|_F$ but increases $\boldsymbol{x}^H \bar{\boldsymbol{R}} \boldsymbol{x}$ by λ. We also note that due to $\|\bar{\boldsymbol{R}} - \boldsymbol{R_1}\|_F$ being monotonically decreasing through the iterations of the method, if the conditions in (47) hold for the solution obtained in any iteration, it will hold for all the iterations afterward.

In the following, we study a suitable diagonal loading of $\boldsymbol{R}$ that ensures meeting the conditions in (47). Next the optimization of the function in (46) is discussed through a separate optimization over the three variables of the problem.

• *Diagonal loading of* $\boldsymbol{R}$: As will be explained later, we can compute $\boldsymbol{Q_1}$ and $\boldsymbol{P_1}$, (hence $\boldsymbol{R_1} = \boldsymbol{Q_1} + \boldsymbol{P_1}$) for any initialization of $\boldsymbol{s}$. In order to guarantee the positive definiteness of $\boldsymbol{R_1}$, define

$$\varepsilon_0 \triangleq \|\bar{\boldsymbol{R}} - \boldsymbol{R_1}\|_F. \tag{49}$$

Then we suggest to diagonally load $\boldsymbol{R}$ with $\lambda > \lambda_0 = -\sigma_n(\boldsymbol{R}) + \varepsilon_0$:

$$\boldsymbol{R} \leftarrow \boldsymbol{R} + \lambda \boldsymbol{I}. \tag{50}$$

• *Optimization with respect to* $\boldsymbol{Q_1}$: We restate the objective function of (46) as

$$\|\boldsymbol{R} - (\boldsymbol{Q_1} + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\|_F = \|\underbrace{(\boldsymbol{R} \odot (\boldsymbol{s}\boldsymbol{s}^H)^* - \boldsymbol{P_1})}_{\boldsymbol{R}_Q} - \boldsymbol{Q_1}\|_F. \tag{51}$$

Given $\boldsymbol{R}_Q$, the partial minimization of (46) with respect to $\boldsymbol{Q_1}$ can be written as

$$\min_{\boldsymbol{Q_1} \in \mathcal{C}_1} \|\boldsymbol{R}_Q - \boldsymbol{Q_1}\|_F. \tag{52}$$

which is equivalent to

$$\begin{aligned} \min_{\boldsymbol{Q_1}, \rho} \quad & \|\boldsymbol{R}_Q - \boldsymbol{Q_1}\|_F \\ \text{s.t.} \quad & \boldsymbol{Q_1}\boldsymbol{1} = \rho \boldsymbol{1}, \\ & \sigma_1(\boldsymbol{Q_1}) = \rho. \end{aligned} \tag{53}$$

In [25], the authors have derived an explicit solution for the optimization problem

$$\begin{aligned} \min_{\boldsymbol{Q_1}} \quad & \|\boldsymbol{R}_Q - \boldsymbol{Q_1}\|_F \\ \text{s.t.} \quad & \boldsymbol{Q_1}\boldsymbol{1} = \rho \boldsymbol{1}. \ (\rho = \text{given}) \end{aligned} \tag{54}$$

The explicit solution of (54) is given by

$$\begin{aligned} \boldsymbol{Q_1}(\rho) &= \rho \boldsymbol{I}_n + \left(\boldsymbol{I}_n - \frac{\boldsymbol{1}_{n\times n}}{n}\right)(\boldsymbol{R}_Q - \rho \boldsymbol{I}_n)\left(\boldsymbol{I}_n - \frac{\boldsymbol{1}_{n\times n}}{n}\right) \\ &= \boldsymbol{R}_Q + \frac{\rho}{n}\boldsymbol{1}_{n\times n} - \frac{2}{n}(\boldsymbol{R}_Q \boldsymbol{1}_{n\times n}) + \frac{1}{n^2}(\boldsymbol{1}_{n\times n}\boldsymbol{R}_Q \boldsymbol{1}_{n\times n}) \end{aligned} \tag{55}$$

[1] i.e., the optimal $\boldsymbol{P_1}$ will be the same as before but with the same diagonal loading.

Note that

$$\boldsymbol{Q_1}(\rho') - \boldsymbol{Q_1}(\rho) = (\rho' - \rho)(\boldsymbol{1}_{n\times 1}/\sqrt{n})(\boldsymbol{1}_{n\times 1}/\sqrt{n})^T \tag{56}$$

which implies that except for the eigenpair $(\boldsymbol{1}_{n\times 1}/\sqrt{n}, \rho)$, all other eigenvalue/vectors are independent of ρ. Let ρ_0 represent the maximal eigenvalue of $\boldsymbol{Q_1}(0)$ corresponding to an eigenvector other than $\boldsymbol{1}_{n\times 1}/\sqrt{n}$. More interesting, the set of the optimal solutions of (54) for different ρ form a line in $\mathbb{C}^{n^2}$ described as in (56). Therefore, (52) is equivalent to

$$\begin{aligned} \min_{\rho} \quad & \|\boldsymbol{R}_Q - \boldsymbol{Q_1}(\rho)\|_F \\ \text{s.t.} \quad & \rho \geq \rho_0. \end{aligned} \tag{57}$$

It follows from (55) that

$$\|\boldsymbol{R}_Q - \boldsymbol{Q_1}(\rho)\|_F^2 = \sum_{k=1}^{n} n \left|\frac{\rho}{n} - \frac{2G_k}{n} + \frac{H}{n^2}\right|^2 \tag{58}$$

where G_k and H are the sum of the kth row and, respectively, the sum of all entries of $\boldsymbol{R}_Q$. The ρ that minimizes (58) is given by

$$\rho = \frac{1}{n}\sum_{k=1}^{n} \Re\left(2G_k - \frac{H}{n}\right) = \frac{H}{n} \tag{59}$$

which implies that the minimizer $\rho = \rho_\star$ of (57) is equal to

$$\rho_\star = \begin{cases} \frac{H}{n} & \frac{H}{n} \geq \rho_0, \\ \rho_0 & \text{otherwise.} \end{cases} \tag{60}$$

Finally, the optimal solution $\boldsymbol{Q_1}$ to (52) is given by

$$\boldsymbol{Q_1} = \boldsymbol{Q_1}(\rho_\star). \tag{61}$$

• *Optimization with respect to* $\boldsymbol{P_1}$: Similar to the previous case, (46) can be rephrased as

$$\min_{\boldsymbol{Q_1} \in \mathcal{C}(\boldsymbol{V}_1)} \|\boldsymbol{R}_P - \boldsymbol{P_1}\|_F \tag{62}$$

where $\boldsymbol{R}_P = \boldsymbol{R} \odot (\boldsymbol{s}\boldsymbol{s}^H)^* - \boldsymbol{Q_1}$. The solution of (62) is simply given by

$$\boldsymbol{P_1}(k,l) = \begin{cases} \boldsymbol{R}'_P(k,l) & \boldsymbol{R}'_P(k,l) \geq 0 \text{ or } k = l, \\ 0 & \text{otherwise} \end{cases} \tag{63}$$

where $\boldsymbol{R}'_P = \Re\{\boldsymbol{R}_P\}$.

• *Optimization with respect to* $\boldsymbol{s}$: Suppose that $\boldsymbol{Q_1}$ and $\boldsymbol{P_1}$ are given and that $\boldsymbol{R_1} = \boldsymbol{Q_1} + \boldsymbol{P_1}$ is a positive definite matrix (see the discussion on this aspect following (46)). Then we have

$$\begin{aligned} &\|\boldsymbol{R} - \boldsymbol{R_1} \odot (\boldsymbol{s}\boldsymbol{s}^H)\|_F^2 \\ &\quad = \|\boldsymbol{R} - \mathbf{Diag}(\boldsymbol{s})\, \boldsymbol{R_1}\, \mathbf{Diag}(\boldsymbol{s}^*)\|_F^2 \\ &\quad = \mathrm{tr}(\boldsymbol{R}^2) + \mathrm{tr}\left(\boldsymbol{R_1}^2\right) - 2\Re\{\mathrm{tr}(\boldsymbol{R}\mathbf{Diag}(\boldsymbol{s})\, \boldsymbol{R_1}\, \mathbf{Diag}(\boldsymbol{s}^*))\}. \end{aligned} \tag{64}$$

Note that only the third term of (64) is a function of $\boldsymbol{s}$. Moreover, it can be verified that [26]

$$\mathrm{tr}(\boldsymbol{R}\, \mathbf{Diag}(\boldsymbol{s})\, \boldsymbol{R_1}\, \mathbf{Diag}(\boldsymbol{s}^*)) = \boldsymbol{s}^H \left(\boldsymbol{R} \odot \boldsymbol{R_1}^T\right) \boldsymbol{s}. \tag{65}$$

As $\boldsymbol{R} \odot \boldsymbol{R_1}^T$ is positive definite, we can employ the power method-like iterations introduced in (28) to decrease the cri-

TABLE I
THE MERIT ALGORITHM

(A) The case of $\alpha_0 = 0$

Step 0: Initialize the variables $\boldsymbol{Q_1}$ and $\boldsymbol{P_1}$ with $\boldsymbol{I}$. Let $\boldsymbol{s}$ be a random vector in Ω^n.
Step 1: Perform the diagonal loading of $\boldsymbol{R}$ as in (49)-(50) (note that this diagonal loading is sufficient to keep $\boldsymbol{R_1} = \boldsymbol{Q_1} + \boldsymbol{P_1}$ always positive definite).
Step 2: Obtain the minimum of (46) with respect to $\boldsymbol{Q_1}$ as in (61).
Step 3: Obtain the minimum of (46) with respect to $\boldsymbol{P_1}$ using (63).
Step 4: Minimize (46) with respect to $\boldsymbol{s}$ using (66).
Step 5: Goto step 2 until a stop criterion is satisfied, e.g. $\|\boldsymbol{R} - (\boldsymbol{Q_1} + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\|_F \leq \epsilon_0$ (or if the number of iterations exceeded a predefined maximum number).

(B) The case of $\alpha_0 > 0$

Step 0: Initialize the variables $(\boldsymbol{s}, \boldsymbol{Q_1}, \boldsymbol{P_1})$ using the results obtained by the optimization of (46) as in Table I-A.
Step 1: Set δ (the step size for increasing α_0 in each iteration). Let δ_0 be the minimal δ to be considered and $\alpha_0 = 0$.
Step 2: Let $\alpha_0^{pre} = \alpha_0$, $\alpha_0^{new} = \alpha_0 + \delta$ and $\boldsymbol{R'} = \boldsymbol{R} + \alpha_0^{new}\boldsymbol{s}\boldsymbol{s}^H$.
Step 3: Solve (67) using the steps 2-5 in Table I-A (particularly step 4 must be applied to (71)).
Step 4: If $\|\boldsymbol{R'} - (\boldsymbol{Q_1} + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\|_F \leq \epsilon_0$ do:

- **Step 4-1**: If $\delta \geq \delta_0$, let $\delta \leftarrow \delta/2$ and initialize (67) with the previously obtained variables $(\boldsymbol{s}, \boldsymbol{Q_1}, \boldsymbol{P_1})$ for $\alpha_0 = \alpha_0^{pre}$. Goto step 2.
- **Step 4-2**: If $\delta < \delta_0$, stop.

Else, let $\alpha_0 = \alpha_0^{new}$ and goto step 2.

terion in (46), i.e., starting from the current $\boldsymbol{s} = \boldsymbol{s}^{(0)}$, a local optimum of the problem can be obtained by the iterations

$$\boldsymbol{s}^{(t+1)} = e^{j\arg((\boldsymbol{R} \odot \boldsymbol{R_1^T})\boldsymbol{s}^{(t)})}. \tag{66}$$

Remark 2: Note that the ability of using more general constraints (e.g., the PAR constraint) in the power method-like iterations means that MERIT can deal with such generalized constraints. This is basically due to the fact that the optimization of the MERIT criterion with respect to $\boldsymbol{s}$ is accomplished via the power method-like iterations. ■

Finally, the proposed algorithmic optimization of (46) based on the above results is summarized in Table I-A.

B. Achieving a Local Optimum of UQP (the Case of $\alpha_0 > 0$)

There exist examples for which the objective function in (46) does not converge to zero. As a result, the proposed method cannot obtain a global optimum of UQP in such cases. However, it is still possible to obtain a local optimum of UQP for some $\alpha_0 > 0$. To do so, we solve the optimization problem,

$$\min_{\boldsymbol{s}\in\Omega, \boldsymbol{Q_1}\in\mathcal{C}_1, \boldsymbol{P_1}\in\mathcal{C}(\boldsymbol{V_1})} \|\boldsymbol{R'} - (\boldsymbol{Q_1} + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\|_F \tag{67}$$

with $\boldsymbol{R'} = \boldsymbol{R} + \alpha_0\boldsymbol{s}\boldsymbol{s}^H$, for increasing α_0. It is worth pointing out that achieving a zero value for the criterion in (67) implies $\boldsymbol{R} + \alpha_0\boldsymbol{s}\boldsymbol{s}^H \in \mathcal{K}(\boldsymbol{s})$. As a result, there exists a non-negative $\boldsymbol{v} \in \mathbb{R}^n$ such that

$$(\boldsymbol{R} + \alpha_0\boldsymbol{s}\boldsymbol{s}^H)\boldsymbol{s} = \boldsymbol{v} \odot \boldsymbol{s}. \tag{68}$$

Consequently,

$$\boldsymbol{R}\boldsymbol{s} = (\boldsymbol{v} - n\alpha_0\boldsymbol{1}) \odot \boldsymbol{s} \tag{69}$$

which implies $\boldsymbol{s}$ is a stationary point of the UQP associated with $\boldsymbol{R}$.

The optimization problem in (67) can be tackled using the same tools as proposed for (46). In particular, note that increasing α_0 decreases (67). To observe this, suppose that the solution $(\boldsymbol{s}, \boldsymbol{Q_1}, \boldsymbol{P_1})$ of (67) is given for an $\alpha_0 \geq 0$. The minimization of (67) with respect to $\boldsymbol{Q_1}$ for $\alpha_0^{\text{new}} = \alpha_0 + \delta$ $(\delta > 0)$ yields $\tilde{\boldsymbol{Q}}_{\boldsymbol{1}} \in \mathcal{C}_{\boldsymbol{1}}$ such that

$$\begin{aligned}&\left\|\boldsymbol{R} + \alpha_0^{\text{new}}\boldsymbol{s}\boldsymbol{s}^H - (\tilde{\boldsymbol{Q}}_{\boldsymbol{1}} + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\right\|_F \\ &\quad\leq \left\|\boldsymbol{R} + \alpha_0^{\text{new}}\boldsymbol{s}\boldsymbol{s}^H - ((\boldsymbol{Q_1} + \delta\boldsymbol{1}\boldsymbol{1}^T) + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\right\|_F \\ &\quad= \|\boldsymbol{R} + \alpha_0\boldsymbol{s}\boldsymbol{s}^H - (\boldsymbol{Q_1} + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\|_F \end{aligned} \tag{70}$$

where $\boldsymbol{Q_1} + \delta\boldsymbol{1}\boldsymbol{1}^T \in \mathcal{C}_{\boldsymbol{1}}$. The optimization of (67) with respect to $\boldsymbol{P_1}$ can be dealt with as before (see (46)) and it leads to a further decrease of the objective function. Furthermore,

$$\begin{aligned}&\|\boldsymbol{R} + \alpha_0\boldsymbol{s}\boldsymbol{s}^H - (\boldsymbol{Q_1} + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\|_F \\ &= \|\boldsymbol{R} + \lambda'\boldsymbol{I} - (\boldsymbol{Q_1} + \boldsymbol{P_1} - \alpha_0\boldsymbol{1}\boldsymbol{1}^T + \lambda'\boldsymbol{I}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\|_F \end{aligned} \tag{71}$$

which implies that a solution $\boldsymbol{s}$ of (67) can be obtained via optimizing (71) with respect to $\boldsymbol{s}$ in a similar way as we described for (46) provided that $\lambda' \geq 0$ is such that $\boldsymbol{Q_1} + \boldsymbol{P_1} - \alpha_0\boldsymbol{1}\boldsymbol{1}^T + \lambda'\boldsymbol{I}$ is positive definite. Finally, note that the obtained solution $(\boldsymbol{s}, \boldsymbol{Q_1}, \boldsymbol{P_1})$ of (46) can be used to initialize the corresponding variables in (67). In effect, the solution of (67) for any α_0 can be used for the initialization of (67) with an increased α_0.

Based on the above discussion and the fact that small values of α_0 are of interest, a bisection approach can be used to obtain α_0. The proposed method for obtaining a local optimum of UQP along with the corresponding α_0 is described in Table I-B. Using the proposed algorithm, the task of finding the minimal α_0 can be accomplished within *finite number of steps*, see Appendix F.

C. Sub-Optimality Analysis

In this sub-section, we show how the proposed method can provide real-time sub-optimality guarantees and bounds during its iterations. Let $\alpha_0 = 0$ (as a result $\boldsymbol{R'} = \boldsymbol{R}$) and define

$$\boldsymbol{E} \triangleq \boldsymbol{R'} - \underbrace{(\boldsymbol{Q_1} + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)}_{\boldsymbol{R_s}} \tag{72}$$

where $\boldsymbol{Q_1} \in \mathcal{C}_{\boldsymbol{1}}$ and $\boldsymbol{P_1} \in \mathcal{C}(\boldsymbol{V_1})$. By construction, the global optimum of the UQP associated with $\boldsymbol{R_s}$ is $\boldsymbol{s}$. We have that

$$\begin{aligned}\max_{\boldsymbol{s'}\in\Omega^n} \boldsymbol{s'}^H\boldsymbol{R}\boldsymbol{s'} &\leq \max_{\boldsymbol{s'}\in\Omega^n} \boldsymbol{s'}^H\boldsymbol{R_s}\boldsymbol{s'} + \max_{\boldsymbol{s'}\in\Omega^n} \boldsymbol{s'}^H\boldsymbol{E}\boldsymbol{s'} \\ &\leq \max_{\boldsymbol{s'}\in\Omega^n} \boldsymbol{s'}^H\boldsymbol{R_s}\boldsymbol{s'} + n\sigma_1(\boldsymbol{E}) \\ &= \boldsymbol{s}^H\boldsymbol{R_s}\boldsymbol{s} + n\sigma_1(\boldsymbol{E}) \end{aligned} \tag{73}$$

Furthermore,

$$\begin{aligned}\max_{\boldsymbol{s'}\in\Omega^n} \boldsymbol{s'}^H\boldsymbol{R}\boldsymbol{s'} &\geq \max_{\boldsymbol{s'}\in\Omega^n} \boldsymbol{s'}^H\boldsymbol{R_s}\boldsymbol{s'} + \min_{\boldsymbol{s'}\in\Omega^n} \boldsymbol{s'}^H\boldsymbol{E}\boldsymbol{s'} \\ &\geq \max_{\boldsymbol{s'}\in\Omega^n} \boldsymbol{s'}^H\boldsymbol{R_s}\boldsymbol{s'} + n\sigma_n(\boldsymbol{E}) \\ &= \boldsymbol{s}^H\boldsymbol{R_s}\boldsymbol{s} + n\sigma_n(\boldsymbol{E}) \end{aligned} \tag{74}$$

As a result, an upper bound and a lower bound on the objective function for the global optimum of (46) can be obtained *at each iteration*. In accordance to what discussed earlier, as

$$|\sigma_1(\boldsymbol{E})| \leq \|\boldsymbol{E}\|_F, \quad |\sigma_n(\boldsymbol{E})| \leq \|\boldsymbol{E}\|_F \tag{75}$$

if $\|\boldsymbol{E}\|_F$ converges to zero we conclude from (73) and (74) that

$$\max_{\boldsymbol{s}' \in \Omega^n} \boldsymbol{s}'^H \boldsymbol{R} \boldsymbol{s}' = \boldsymbol{s}^H \boldsymbol{R}_{\boldsymbol{s}} \boldsymbol{s} = \boldsymbol{s}^H \boldsymbol{R} \boldsymbol{s} \tag{76}$$

and hence $\boldsymbol{s}$ is the global optimum of the UQP associated with $\boldsymbol{R}$ (i.e., a sub-optimality guarantee of $\gamma = 1$ is achieved).

Next, suppose that we have to increase α_0 in order to obtain the convergence of $\|\boldsymbol{E}\|_F$ to zero. In such a case, we have that $\boldsymbol{R} = \boldsymbol{R}_{\boldsymbol{s}} - \alpha_0 \boldsymbol{s}\boldsymbol{s}^H$ and as a result, $\max_{\boldsymbol{s}' \in \Omega^n} \boldsymbol{s}'^H \boldsymbol{R}_{\boldsymbol{s}} \boldsymbol{s}' - \alpha_0 n^2 \leq \max_{\boldsymbol{s}' \in \Omega^n} \boldsymbol{s}'^H \boldsymbol{R} \boldsymbol{s}' \leq \max_{\boldsymbol{s}' \in \Omega^n} \boldsymbol{s}'^H \boldsymbol{R}_{\boldsymbol{s}} \boldsymbol{s}'$ or equivalently,

$$\boldsymbol{s}^H \boldsymbol{R}_{\boldsymbol{s}} \boldsymbol{s} - \alpha_0 n^2 \leq \max_{\boldsymbol{s}' \in \Omega^n} \boldsymbol{s}'^H \boldsymbol{R} \boldsymbol{s}' \leq \boldsymbol{s}^H \boldsymbol{R}_{\boldsymbol{s}} \boldsymbol{s}. \tag{77}$$

The provided case-dependent sub-optimality guarantee is thus given by

$$\gamma = \frac{\boldsymbol{s}^H \boldsymbol{R} \boldsymbol{s}}{\boldsymbol{s}^H \boldsymbol{R}_{\boldsymbol{s}} \boldsymbol{s}} = 1 - \frac{\alpha_0 n^2}{\boldsymbol{s}^H \boldsymbol{R}_{\boldsymbol{s}} \boldsymbol{s}} = \frac{\boldsymbol{s}^H \boldsymbol{R} \boldsymbol{s}}{\boldsymbol{s}^H \boldsymbol{R} \boldsymbol{s} + \alpha_0 n^2}. \tag{78}$$

The following section provides empirical evidence to the fact that (78) can yield tighter sub-optimality guarantees than the currently known approximation guarantee of $\pi/4$ for SDR.

VI. Numerical Examples

In order to examine the performance of the proposed method, several numerical examples will be presented. Random Hermitian matrices $\boldsymbol{R}$ are generated using the formula

$$\boldsymbol{R} = \sum_{k=1}^{n} \boldsymbol{x}_k \boldsymbol{x}_k^H \tag{79}$$

where $\{\boldsymbol{x}_k\}$ are random vectors in $\mathbb{C}^n$ whose real-part and imaginary-part elements are i.i.d. with a standard Gaussian distribution $\mathcal{N}(0,1)$. In all cases, we stopped the iterations when $\|\boldsymbol{E}\|_F \leq 10^{-9}$.

We use the MERIT algorithm to solve the UQP for a random positive definite matrix of size $n = 16$. The obtained values of the UQP objective for the true matrix $\boldsymbol{R}$ and the approximated matrix $\boldsymbol{R}_{\boldsymbol{s}}$ as well as the sub-optimality bounds (derived in (73) and (74)) are depicted in Fig. 2 versus the iteration number. In this example, a sub-optimality guarantee of $\gamma = 1$ is achieved which implies that the method has successfully obtained the global optimum of the considered UQP. A computational time of 3.653 sec was required to accomplish the task on a standard PC. For the sake of comparison, we also use the power method-like iterations discussed in Section III, and MERIT, as well as the curvilinear search of [9] with Barzilai-Borwein (BB) step size, to solve an UQP ($n = 10$) based on the same initialization. The resultant UQP objectives along with required times (in sec) versus iteration number are plotted in Fig. 3. It can be observed that the power method-like iterations approximate the UQP solution much faster than the curvilinear search of [9]. On the other hand, both methods are much faster than MERIT. This type of behavior, which is not unexpected, is due

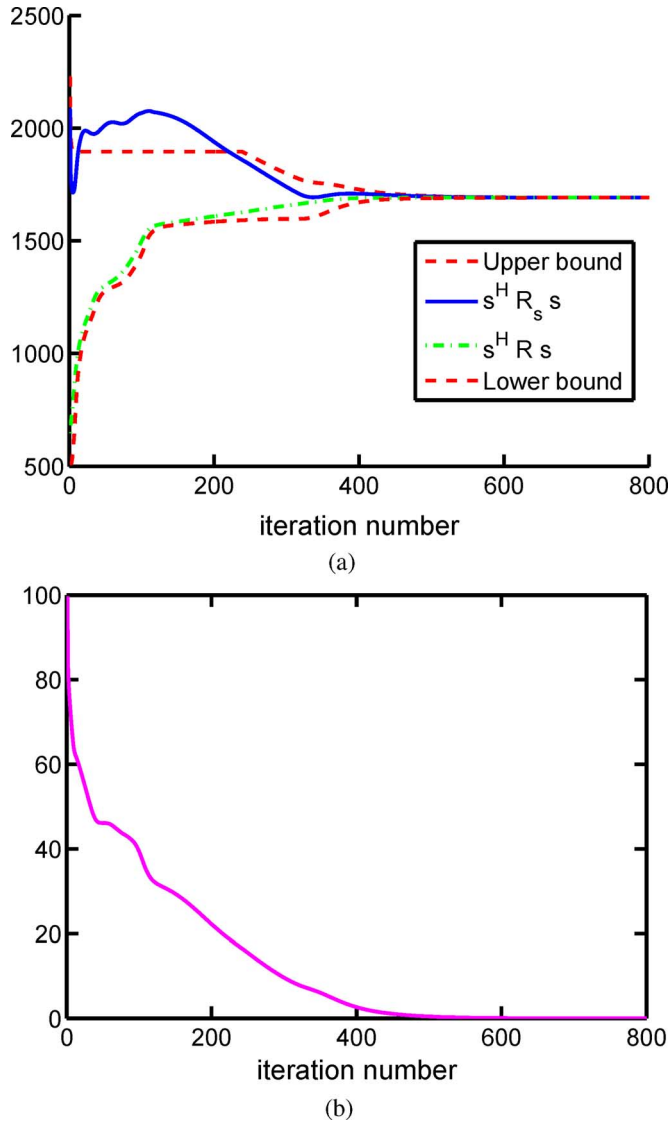


Fig. 2. Different metrics versus the iteration number for an UQP solved by MERIT. (a) the UQP objective corresponding to the true matrix $\boldsymbol{R}$, the approximated matrix $\boldsymbol{R}_{\boldsymbol{s}}$ and also the upper/lower bounds at each iteration. The sub-optimality bounds are updated using (73)–(74). (b) the criterion $\|\boldsymbol{E}\|_F = \|\boldsymbol{R} - \boldsymbol{R}_s\|_F$ (it reaches values which are practically zero).

to the fact that MERIT is not designed solely for local optimization; indeed, MERIT relies on a considerable over-parametrization in its formulation which is the cost paid for easily derivable sub-optimality guarantees. In general, one may employ the power method-like iterations to obtain a fast approximation of the UQP solution (e.g., by using several initializations), whereas for obtaining sub-optimality guarantees one can resort to MERIT.

Next, we approximate the UQP solutions for 20 full-rank random positive definite matrices of sizes $n \in \{8, 16, 32, 64\}$. Inspired by [12] and [27], we also consider rank-deficient matrices $\boldsymbol{R} = \sum_{k=1}^{d} \boldsymbol{x}_k \boldsymbol{x}_k^H$ where $\{\boldsymbol{x}_k\}$ are as in (79), but $d \ll n$. The performance of MERIT for different values of d is shown in Table II. Interestingly, the solution of UQP for rank-deficient matrices appears to be more efficiently obtained than for full-rank matrices. We also employ SDR [19] to solve the same UQPs. Note also that given the solutions obtained by MERIT

TABLE II
COMPARISON OF THE PERFORMANCE OF MERIT (SEE TABLE I) AND SDR [19] WHEN SOLVING THE UQP FOR 20 RANDOM POSITIVE DEFINITE MATRICES OF DIFFERENT SIZES n AND RANKS d

n	Rank (d)	#problems for which $\gamma = 1$	Average γ	Minimum γ	Average MERIT CPU time (sec)	Average SDR time / Average MERIT time
8	*2*	17	0.9841	0.8184	0.13	1.08
	8	16	0.9912	0.9117	0.69	0.81
16	*2*	15	0.9789	0.8301	1.06	2.08
	4	13	0.9773	0.8692	1.58	0.95
	16	4	0.9610	0.8693	3.54	0.92
32	*2*	9	0.9536	0.8190	47.04	4.79
	6	4	0.9077	0.8106	55.59	2.44
	32	2	0.9031	0.8021	94.90	1.12
64	*2*	3	0.8893	0.8177	406.56	4.34
	8	1	0.8567	0.7727	560.35	2.07
	64	0	0.8369	0.7811	1017.69	1.85

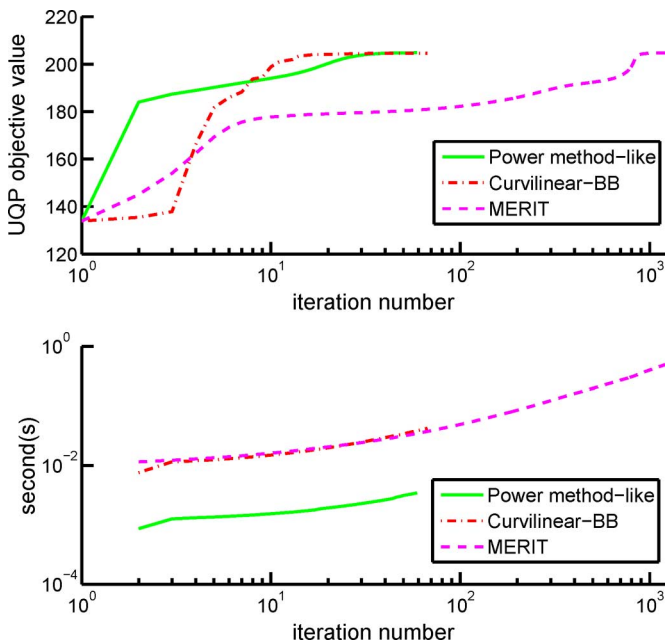


Fig. 3. A comparison of power method-like iterations, the curvilinear search of [9] with Barzilai-Borwein (BB) step size, and MERIT: (top) the UQP objective; (bottom) the required time for solving an UQP ($n = 10$) with same initialization. The computation times for the three methods were 3.44×10^{-3}, 4.22×10^{-2}, and 4.89×10^{-1} (all in sec), respectively.

and SDR as well as the sub-optimality guarantee of MERIT, a case-dependent sub-optimality guarantee for SDR can be computed as

$$\gamma_{\rm SDR} \triangleq \gamma_{\rm MERIT} \left(\frac{v_{\rm SDR}}{v_{\rm MERIT}} \right). \tag{80}$$

This can be used to examine the goodness of the solutions obtained by SDR. In this example, we continue the randomization procedure of SDR until reaching the same UQP objective as for MERIT. The results can be found in Table II. The results imply that, although the average SDR time is less than MERIT in some cases, the average MERIT time appears to outperform that of SDR for larger dimensions n or lower matrix ranks d.

As discussed earlier, the UQP formulation occurs in different code design scenarios. An interesting code design problem arises when synthesizing waveforms that have good resolution properties in range and Doppler [3]–[5], [28]–[31]. In the following, we consider the design of a thumbtack CAF (see the definitions in Section I-A):

$$d(\tau, f) = \begin{cases} n & (\tau, f) = (0, 0), \\ 0 & \text{otherwise.} \end{cases} \tag{81}$$

Suppose $n = 53$, let T be the time duration of the total waveform, and let $t_p = T/n$ represent the time duration of each sub-pulse. Define the weighting function as

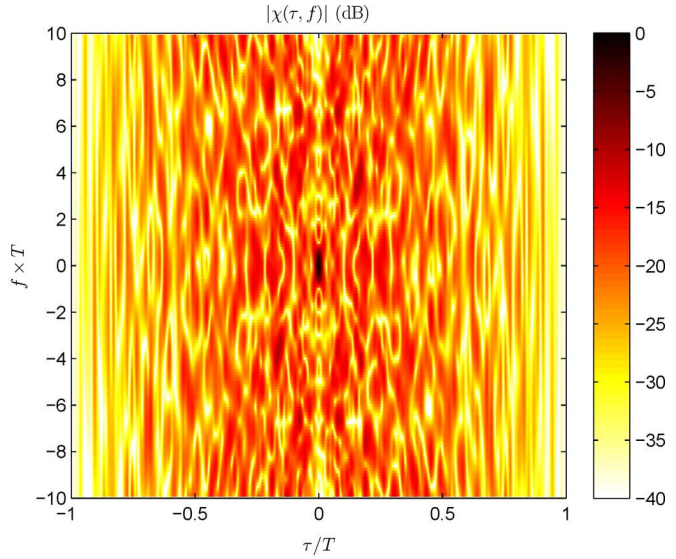
$$w(\tau, f) = \begin{cases} 1 & (\tau, f) \in \Psi \backslash \Psi_{ml}, \\ 0 & \text{otherwise,} \end{cases} \tag{82}$$

where $\Psi = [-10t_p, 10t_p] \times [-2/T, 2/T]$ is the region of interest and $\Psi_{ml} = ([-t_p, t_p] \backslash \{0\}) \times ([-1/T, 1/T] \backslash \{0\})$ is the mainlobe area which is excluded due to the sharp changes near the origin of $d(\tau, f)$. Note that the time delay τ and the Doppler frequency f are typically normalized by T and $1/T$, respectively, and as a result the value of t_p can be chosen freely without changing the performance of CAF design. The synthesis of the desired CAF is accomplished via the cyclic minimization of (8) with respect to $\boldsymbol{x}$ and $\boldsymbol{y}$ (see Section I-A). In particular, we use MERIT to obtain a unimodular $\boldsymbol{x}$ in each iteration. A Björck code is used to initialize both vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. The Björck code of length $n = p$ (where p is a prime number for which $p \equiv 1 (\mathrm{mod} 4)$) is given by $\boldsymbol{b}(k) = e^{j(\frac{k}{p}) \arccos(1/(1+\sqrt{p}))}$, $0 \leq k < p$, with $(\frac{k}{p})$ denoting the Legendre symbol. Fig. 4 depicts the normalized CAF modulus of the Björck code (i.e., the initial CAF) and the obtained CAF using the UQP formulation in (11) and the proposed method. Despite the fact that designing CAF with a unimodular transmit vector $\boldsymbol{x}$ is a rather difficult problem, MERIT is able to efficiently suppress the CAF sidelobes in the region of interest.

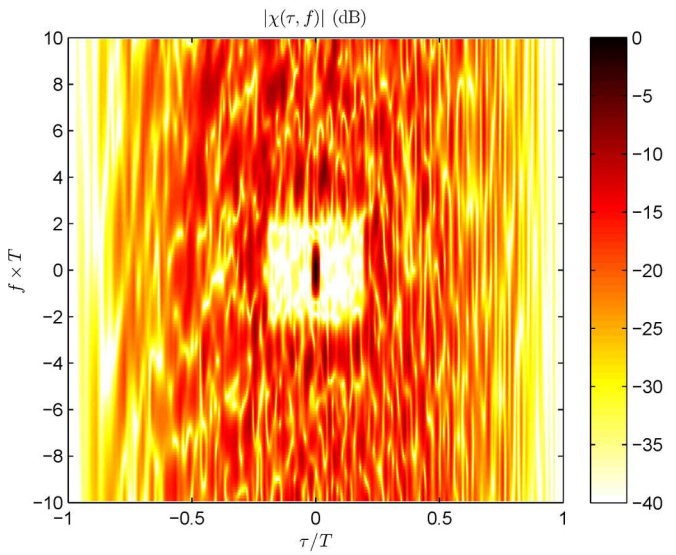
VII. CONCLUDING REMARKS

A computational approach to the NP-hard problem of optimizing a quadratic form over the unimodular vector set (called UQP) has been introduced. The main results can be summarized as follows:

- Power method-like iterations were devised for local optimization. The proposed method was shown to be useful not only for the quadratic optimization over unimodular codes but also for some other types of code constraints. The particular example of PAR constrained code design was discussed in some detail.



(a)



(b)

Fig. 4. The normalized CAF modulus for (a) the Björck code of length $n = 53$ (i.e., the initial CAF), and (b) the UQP formulation in (11) and MERIT.

- It was shown that the set of matrices $\mathcal{K}(\boldsymbol{s})$ leading to the same solution $\boldsymbol{s}$ as the global optimum of UQP is a convex cone. An one-to-one mapping between any two such convex cones was introduced and an approximate characterization of $\mathcal{K}(\boldsymbol{s})$ was proposed.
- Using the approximate characterization of $\mathcal{K}(\boldsymbol{s})$, an iterative approach (called MERIT) to the UQP was proposed. It was shown that MERIT provides real-time case-dependent sub-optimality guarantees during its iterations. The available numerical evidence shows that the sub-optimality guarantees obtained by MERIT may be better than the currently known approximation guarantee of $\pi/4$ for SDR.

We note that more rigorous efficiency assessments of the method would be useful. It is clear that $\mathcal{C}(\boldsymbol{V_s}) \oplus \mathcal{C}_{\boldsymbol{s}} \subset \mathcal{K}(\boldsymbol{s})$. A possible approach would be to determine how large is the part of $\mathcal{K}(\boldsymbol{s})$ that is "covered" by $\mathcal{C}(\boldsymbol{V_s}) \oplus \mathcal{C}_{\boldsymbol{s}}$; a research problem which is left for future work. Furthermore, a study of m-UQP using the ideas in this paper will be the subject of separate paper.

APPENDIX

A. Proof of Theorem 1

In order to verify the first part of the theorem, consider any two matrices $\boldsymbol{R}_1, \boldsymbol{R}_2 \in \mathcal{K}(\tilde{\boldsymbol{s}})$. For any two non-negative scalars γ_1, γ_2 we have that

$$\boldsymbol{s}^H(\gamma_1 \boldsymbol{R}_1 + \gamma_2 \boldsymbol{R}_2)\boldsymbol{s} = \gamma_1 \boldsymbol{s}^H \boldsymbol{R}_1 \boldsymbol{s} + \gamma_2 \boldsymbol{s}^H \boldsymbol{R}_2 \boldsymbol{s}. \tag{83}$$

Clearly, if some $\boldsymbol{s} = \tilde{\boldsymbol{s}}$ is the global maximizer of both $\boldsymbol{s}^H \boldsymbol{R}_1 \boldsymbol{s}$ and $\boldsymbol{s}^H \boldsymbol{R}_2 \boldsymbol{s}$ then it is the global maximizer of $\boldsymbol{s}^H(\gamma_1 \boldsymbol{R}_1 + \gamma_2 \boldsymbol{R}_2)\boldsymbol{s}$ which implies $\gamma_1 \boldsymbol{R}_1 + \gamma_2 \boldsymbol{R}_2 \in \mathcal{K}(\tilde{\boldsymbol{s}})$.

The second part of the theorem can be shown noting that

$$\begin{aligned} \boldsymbol{s}_2^H \left(\boldsymbol{R} \odot \left(\boldsymbol{s}_0 \boldsymbol{s}_0^H\right)\right) \boldsymbol{s}_2 &= (\boldsymbol{s}_0^* \odot \boldsymbol{s}_2)^H \boldsymbol{R} (\boldsymbol{s}_0^* \odot \boldsymbol{s}_2) \\ &= \boldsymbol{s}_1^H \boldsymbol{R} \boldsymbol{s}_1 \end{aligned} \tag{84}$$

for all $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \Omega^n$ and $\boldsymbol{s}_0 = \boldsymbol{s}_1^* \odot \boldsymbol{s}_2$. Therefore, if $\boldsymbol{R} \in \mathcal{K}(\tilde{\boldsymbol{s}}_1)$ then $\boldsymbol{R} \odot (\tilde{\boldsymbol{s}}_0 \tilde{\boldsymbol{s}}_0^H) \in \mathcal{K}(\tilde{\boldsymbol{s}}_2)$ (for $\tilde{\boldsymbol{s}}_0 = \tilde{\boldsymbol{s}}_1^* \odot \tilde{\boldsymbol{s}}_2$) and vice versa.

B. Characterization of the Stationary Points and Optima of UQP

Let $\boldsymbol{s} = (e^{j\phi_1}, \ldots, e^{j\phi_n})^T$ and note that

$$\begin{aligned} L = \boldsymbol{s}^H \boldsymbol{R} \boldsymbol{s} &= \sum_{k,l} \boldsymbol{R}(k,l) e^{j(\phi_l - \phi_k)} \\ &= \frac{1}{2} \left(\sum_{k,l} \boldsymbol{R}(k,l) e^{j(\phi_l - \phi_k)} + \sum_{k,l} \boldsymbol{R}(l,k) e^{j(\phi_k - \phi_l)} \right). \end{aligned} \tag{85}$$

To obtain the stationary points of UQP (associated with $\boldsymbol{R}$) one can write the following partial derivative equations for all $1 \leq k_0 \leq n$:

$$\begin{aligned} \frac{\partial L}{\partial \phi_{k_0}} &= \frac{1}{2} \left(-j \sum_l \boldsymbol{R}(k_0, l) e^{j(\phi_l - \phi_{k_0})} \right. \\ &\qquad \left. + j \sum_l \boldsymbol{R}(l, k_0) e^{j(\phi_{k_0} - \phi_l)} \right) \\ &= \Im \left\{ e^{-j\phi_{k_0}} \left(\sum_l \boldsymbol{R}(k_0, l) e^{j\phi_l} \right) \right\} = 0 \end{aligned} \tag{86}$$

which implies that there exist $v_{k_0} \in \mathbb{R}$ such that

$$\sum_l \boldsymbol{R}(k_0, l) e^{j\phi_l} = v_{k_0} e^{j\phi_{k_0}}. \tag{87}$$

Considering the above set of equations for all $1 \leq k_0 \leq n$ yields the characterization of the stationary points of L as

$$\boldsymbol{R}\check{\boldsymbol{s}} = \boldsymbol{v} \odot \check{\boldsymbol{s}} \tag{88}$$

where $\boldsymbol{v} \in \mathbb{R}^n$. Based on the latter characterization of the stationary points, we study the optima of UQP by employing the second derivatives of L. For any $l_0 \neq k_0$ we have that

$$\begin{aligned}\frac{\partial^2 L}{\partial\phi_{k_0}\partial\phi_{l_0}} &= \frac{1}{2}\Big(-j(j)\boldsymbol{R}(k_0,l_0)e^{j(\phi_{l_0}-\phi_{k_0})} \\ &\quad +j(-j)\boldsymbol{R}(l_0,k_0)e^{j(\phi_{k_0}-\phi_{l_0})}\Big) \\ &= \Re\left\{\boldsymbol{R}(k_0,l_0)e^{j(\phi_{l_0}-\phi_{k_0})}\right\}. \end{aligned} \tag{89}$$

For $l_0 = k_0$ we can write

$$\begin{aligned}\frac{\partial^2 L}{\partial\phi_{k_0}^2} &= \frac{1}{2}\left(-j(-j)\sum_{l\neq k_0}\boldsymbol{R}(k_0,l)e^{j(\phi_l-\phi_{k_0})}\right. \\ &\quad \left.+j(j)\sum_{l\neq k_0}\boldsymbol{R}(l,k_0)e^{j(\phi_{k_0}-\phi_l)}\right) \\ &= \Re\left\{\boldsymbol{R}(k_0,k_0) - v_{k_0}\right\}. \end{aligned} \tag{90}$$

Therefore, the Hessian matrix associated with L is given by $\boldsymbol{H} = \Re\{\boldsymbol{R}\odot(\boldsymbol{s}\boldsymbol{s}^H)^* - \boldsymbol{V}\}$ where $\boldsymbol{V} = \mathbf{Diag}(\boldsymbol{v})$. As a direct consequence, $\boldsymbol{s}$ is a local maximum of UQP iff $\boldsymbol{H} \leq 0$, or equivalently $\boldsymbol{V} \geq \boldsymbol{R}$.

C. Proof of Theorem 4

It is worthwhile to observe that the convergence rate of $\{\boldsymbol{R}^{(t)}\}$ is not dependent on the problem dimension (n), as each entry of $\{\boldsymbol{R}^{(t)}\}$ is treated independently from the other entries (i.e., all the operations are element-wise). Therefore, without loss of generality we study the convergence of one entry (say $\{\boldsymbol{R}^{(t)}(k,l)\} = \{r_t e^{j\theta_t}\}$) in the following.

Note that in cases for which $|\theta_t - (\phi_k - \phi_l)| > \pi/2$, the next element of the sequence $\{r_t e^{j\theta_t}\}$ can be written as

$$r_{t+1}e^{j\theta_{t+1}} = r_t e^{j\theta_t} + \rho e^{j(\phi_k-\phi_l)} \tag{91}$$

which implies that the proposed operation tends to make θ_t closer to $(\phi_k - \phi_l)$ in each iteration, and finally puts θ_t within the $\pi/2$ distance from $(\phi_k - \phi_l)$.

Let us suppose that $|\theta_0 - (\phi_k - \phi_l)| > \pi/2$, and that the latter phase criterion remains satisfied for all θ_t, $t < T$. We have that

$$r_T e^{j\theta_T} = r_0 e^{j\theta_0} + T\rho e^{j(\phi_k-\phi_l)} \tag{92}$$

which yields

$$r_T\cos(\theta_T - (\phi_k-\phi_l)) = r_0\cos(\theta_0-(\phi_k-\phi_l)) + T\rho. \tag{93}$$

Therefore it takes only $T = \lceil -r_0\cos(\theta_0-(\phi_k-\phi_l))/\rho\rceil = 1$ iteration for θ_t to stand within the $\pi/2$ distance from $(\phi_k - \phi_l)$.

Now, suppose that $|\theta_0 - (\phi_k - \phi_l)| \leq \pi/2$. For every $t \geq 1$ we can write that

$$\begin{aligned} r_{t+1}e^{j\theta_{t+1}} &= r_t e^{j\theta_t} + \rho e^{j(\phi_k-\phi_l)} \\ &\quad - r_t\cos(\theta_t-(\phi_k-\phi_l))e^{j(\phi_k-\phi_l)} \\ &= e^{j(\phi_k-\phi_l)}(\rho + jr_t\sin(\theta_t-(\phi_k-\phi_l))). \end{aligned} \tag{94}$$

Let $\delta_{t+1} = r_{t+1}e^{j\theta_{t+1}} - r_t e^{j\theta_t}$. The first equality in (94) implies that

$$\delta_{t+1} = e^{j(\phi_k-\phi_l)}(\rho - r_t\cos(\theta_t-(\phi_k-\phi_l))). \tag{95}$$

On the other hand, the second equality in (94) implies that

$$\begin{aligned}\delta_{t+1} &= e^{j(\phi_k-\phi_l)}(\rho + jr_t\sin(\theta_t-(\phi_k-\phi_l))) \\ &\quad - e^{j(\phi_k-\phi_l)}(\rho + jr_{t-1}\sin(\theta_{t-1}-(\phi_k-\phi_l))) \\ &= je^{j(\phi_k-\phi_l)}(r_t\sin(\theta_t-(\phi_k-\phi_l)) \\ &\quad - r_{t-1}\sin(\theta_{t-1}-(\phi_k-\phi_l))) \end{aligned} \tag{96}$$

for all $t \geq 1$. Note that in (95) and (96), δ_{t+1} is a complex number having different phases. We conclude

$$\delta_{t+1} = 0, \quad \forall\, t \geq 1 \tag{97}$$

which shows that the sequence $\{r_t e^{j\theta_t}\}$ is convergent in one iteration. In sum, every entry of the matrix R will converge in at most two iterations (i.e., at most one to achieve a phase value within the $\pi/2$ distance from $(\phi_k - \phi_l)$, and one iteration thereafter).

D. Proof of Theorem 5

We use the same notations as in the proof of Theorem 4. If $|\theta_0 - (\phi_k - \phi_l)| \leq \pi/2$ then

$$\begin{aligned} r_2e^{j\theta_2} &= r_1e^{j\theta_1} \\ &= r_0e^{j\theta_0} + \rho e^{j(\phi_k-\phi_l)} \\ &\quad - r_0\cos(\theta_0-(\phi_k-\phi_l))e^{j(\phi_k-\phi_l)}. \end{aligned} \tag{98}$$

On the other hand, if $|\theta_0 - (\phi_k - \phi_l)| > \pi/2$ we have that $r_1e^{j\theta_1} = r_0e^{j\theta_0} + \rho e^{j(\phi_k-\phi_l)}$. As a result, $r_1\cos(\theta_1 - (\phi_k - \phi_l)) = \rho + r_0\cos(\theta_0 - (\phi_k - \phi_l))$ which implies

$$\begin{aligned} r_2e^{j\theta_2} &= r_1e^{j\theta_1} + \rho e^{j(\phi_k-\phi_l)} \\ &\quad - r_1\cos(\theta_1-(\phi_k-\phi_l))e^{j(\phi_k-\phi_l)} \\ &= \underbrace{r_0e^{j\theta_0} + \rho e^{j(\phi_k-\phi_l)}}_{r_1e^{j\theta_1}} + \rho e^{j(\phi_k-\phi_l)} \\ &\quad - \underbrace{(\rho + r_0\cos(\theta_0-(\phi_k-\phi_l)))}_{r_1\cos(\theta_1-(\phi_k-\phi_l))}e^{j(\phi_k-\phi_l)} \\ &= r_0e^{j\theta_0} + \rho e^{j(\phi_k-\phi_l)} \\ &\quad - r_0\cos(\theta_0-(\phi_k-\phi_l))e^{j(\phi_k-\phi_l)}. \end{aligned} \tag{99}$$

Now, it is easy to verify that (41) follows directly from (98) and (99).

E. Proof of Theorem 6

If $\boldsymbol{s}$ is a stable point of UQP associated with $\boldsymbol{R}^{(0)} = \boldsymbol{R}$ then we have that $\arg(\boldsymbol{s}) = \arg(\boldsymbol{R}\boldsymbol{s})$. Let $\boldsymbol{R}\boldsymbol{s} = \boldsymbol{v}\odot\boldsymbol{s}$ where $\boldsymbol{v}$ is a non-negative real-valued vector in $\mathbb{R}^n$. It follows that

$$\boldsymbol{v}(k)e^{j\phi_k} = \sum_{l=1}^{n}|\boldsymbol{R}(k,l)|e^{j\theta_{k,l}}e^{j\phi_l} \tag{100}$$

or equivalently

$$\boldsymbol{v}(k) = \sum_{l=1}^{n} |\boldsymbol{R}(k,l)| e^{j(\theta_{k,l}-(\phi_k-\phi_l))} \tag{101}$$

which implies that

$$\begin{cases} \sum_{l=1}^{n} |\boldsymbol{R}(k,l)| \cos(\theta_{k,l}-(\phi_k-\phi_l)) \geq 0 \\ \sum_{l=1}^{n} |\boldsymbol{R}(k,l)| \sin(\theta_{k,l}-(\phi_k-\phi_l)) = 0 \end{cases} \tag{102}$$

for all $1 \leq k \leq n$. Now, note that the recursive formula of the sequence $\{\boldsymbol{R}^{(t)}\}$ can be rewritten as

$$\boldsymbol{R}^{(t+1)} = \boldsymbol{R}^{(t)} - \mathbf{Diag}(\boldsymbol{s}) \left(\boldsymbol{R}_{+}^{\boldsymbol{s}(t)} - \rho \mathbf{1}_{n\times n} \right) \mathbf{Diag}(\boldsymbol{s}^*) \tag{103}$$

and as a result,

$$\boldsymbol{R}^{(t+1)}\boldsymbol{s} = \boldsymbol{R}^{(t)}\boldsymbol{s} - \mathbf{Diag}(\boldsymbol{s}) \left(\boldsymbol{R}_{+}^{\boldsymbol{s}(t)} - \rho \mathbf{1}_{n\times n} \right) \mathbf{1}_{n\times 1}. \tag{104}$$

It follows from (104) that if $\boldsymbol{s}$ is a stable point of the UQP associated with $\boldsymbol{R}^{(t)}$ (which implies the existence of non-negative real-valued vector $\boldsymbol{v}^{(t)}$ such that $\boldsymbol{R}^{(t)}\boldsymbol{s} = \boldsymbol{v}^{(t)} \odot \boldsymbol{s}$), then there exists $\boldsymbol{v}^{(t+1)} \in \mathbb{R}^n$ for which $\boldsymbol{R}^{(t+1)}\boldsymbol{s} = \boldsymbol{v}^{(t+1)} \odot \boldsymbol{s}$ and therefore,

$$\begin{aligned} \boldsymbol{v}^{(t+1)}(k)\, e^{j\phi_k} = & \sum_{l=1}^{n} \left| \boldsymbol{R}^{(t)}(k,l) \right| e^{j\theta_{k,l}} e^{j\phi_l} \\ & - \left(\left(\sum_{l=1}^{n} \boldsymbol{R}_{+}^{\boldsymbol{s}\,(t)}(k,l) \right) - n\rho \right) e^{j\phi_k}. \end{aligned} \tag{105}$$

Equation (105) can be rewritten as

$$\begin{aligned} \boldsymbol{v}^{(t+1)}(k) = & \sum_{l=1}^{n} \left| \boldsymbol{R}^{(t)}(k,l) \right| e^{j(\theta_{k,l}-(\phi_k-\phi_l))} \\ & - \left(\sum_{l=1}^{n} \boldsymbol{R}_{+}^{\boldsymbol{s}\,(t)}(k,l) \right) + n\rho \end{aligned} \tag{106}$$

As indicated earlier, $\boldsymbol{s}$ being a stable point for $\boldsymbol{R}^{(0)}$ assures that the imaginary part of (106) is zero. To show that $\boldsymbol{s}$ is a stable point of the UQP associated with $\boldsymbol{R}^{(t+1)}$, we only need to verify that $\boldsymbol{v}^{(t+1)}(k) \geq 0$:

$$\begin{aligned} \boldsymbol{v}^{(t+1)}(k) = & \sum_{l=1}^{n} \left| \boldsymbol{R}^{(t)}(k,l) \right| \cos(\theta_{k,l}-(\phi_k-\phi_l)) \\ & - \left(\sum_{l=1}^{n} \boldsymbol{R}_{+}^{\boldsymbol{s}\,(t)}(k,l) \right) + n\rho \\ = & \, n\rho \\ & + \sum_{l:(k,l)\notin\Theta} \left| \boldsymbol{R}^{(t)}(k,l) \right| \cos(\theta_{k,l}-(\phi_k-\phi_l)) \end{aligned} \tag{107}$$

Now note that the positivity of $\boldsymbol{v}^{(t+1)}(k)$ is concluded from (38). In particular, based on the discussions in the proof of Theorem 4, for $t=1$, there is no $\theta_{k,l}$ such that $|\theta_{k,l}-(\phi_k-\phi_l)| \geq \pi/2$ and therefore $\boldsymbol{v}^{(2)}(k) = n\rho$ for all $1 \leq k \leq n$. As a result,

$$\boldsymbol{R}^{(2)}\boldsymbol{s} = n\rho\boldsymbol{s} \tag{108}$$

which implies that $\boldsymbol{s}$ is an eigenvector of $\boldsymbol{R}^{(2)}$ corresponding to the eigenvalue $n\rho$.

F. Finding the Minimal α_0 in Table I-B (Case of $\alpha_0 > 0$) Requires a Finite Number of Steps

The results of Section IV provide a theoretical upper bound on the minimal α_0 for which $\boldsymbol{R} + \alpha_0 \boldsymbol{s}\boldsymbol{s}^H \in \mathcal{C}(\boldsymbol{V_s}) \oplus \mathcal{C}_{\boldsymbol{s}}$. Note that $\mathcal{C}(\boldsymbol{V_s}) \oplus \mathcal{C}_{\boldsymbol{s}}$ is a convex cone, implying that any such α_0 (for which $\boldsymbol{R} + \alpha_0 \boldsymbol{s}\boldsymbol{s}^H \in \mathcal{C}(\boldsymbol{V_s}) \oplus \mathcal{C}_{\boldsymbol{s}}$) would easily set the objective function of (67) to zero. Equation (43) suggests that any $\alpha_0 \geq 2\rho$ can serve as the upper bound for the minimal α_0. Theorem 7 suggests that any $\rho \geq \mu/n$, where μ is the largest eigenvalue of $\boldsymbol{R}^{(2)}$ belonging to an eigenvector other than $\boldsymbol{s}$, can be used to construct such an upper bound on the minimal α_0. Using the results of Theorems 5 and 6 along with (43) implies that it is sufficient to consider

$$\begin{aligned} \rho &= \sigma_1 \left(\boldsymbol{R}^{(2)}(0) \right) \Big/ n \\ &= \sigma_1 \left(\boldsymbol{R} - \left(\boldsymbol{R}_{+}^{\boldsymbol{s}(0)} + \boldsymbol{R}_{+}^{\boldsymbol{s}(1)} \right) \odot (\boldsymbol{s}\boldsymbol{s}^H) \right) \Big/ n \\ &\leq \frac{3}{n} \|\boldsymbol{R}\|_F \end{aligned} \tag{109}$$

due to the definition of $\boldsymbol{R}_{+}^{\boldsymbol{s}}$ in (37). As a result, it is sufficient to consider

$$\alpha_0 = 2\rho \leq \frac{6}{n} \|\boldsymbol{R}\|_F \tag{110}$$

as an upper bound on the values of α_0 for which the objective function of (67) attains zero. It was shown in (70) that the objective function of (67) is monotonically decreasing with respect to α_0. Considering a step size δ for increasing α_0, it takes at most

$$\eta_1 = \left\lceil \left(\frac{6}{n} \|\boldsymbol{R}\|_F \right) \Big/ \delta \right\rceil \tag{111}$$

steps to achieve $f(\alpha_0^{\text{new}}) \triangleq \|\boldsymbol{R}' - (\boldsymbol{Q_1} + \boldsymbol{P_1}) \odot (\boldsymbol{s}\boldsymbol{s}^H)\|_F \leq \epsilon_0$ in Step 4 of Table I-B, where the bisection procedure starts. In each bisection, the step size δ will be divided by 2, until reaching a priori given precision (δ_0) of the obtained α_0. More precisely, let us suppose $(\alpha_0^{\text{pre}}, \alpha_0^{\text{new}}) = (\alpha_0, \alpha_0 + \delta)$ where $f(\alpha_0^{\text{new}}) = f(\alpha_0 + \delta) \leq \epsilon_0$. At the next step, the bisection approach uses the new pair $(\alpha_0^{\text{pre}}, \alpha_0^{\text{new}}) = (\alpha_0, \alpha_0 + \delta/2)$ to check whether $f(\alpha_0 + \delta/2) \leq \epsilon_0$. If $f(\alpha_0 + \delta/2) \leq \epsilon_0$ then the bisection approach will be recursively applied for $(\alpha_0^{\text{pre}}, \alpha_0^{\text{new}}) = (\alpha_0, \alpha_0 + \delta/2)$. Otherwise, the algorithm considers $(\alpha_0^{\text{pre}}, \alpha_0^{\text{new}}) = (\alpha_0 + \delta/2, \alpha_0 + \delta)$ as the new candidate for applying the bisection procedure. Therefore, the number of steps required to obtain α_0 with a fixed precision δ_0 is given by

$$\eta_2 = \lceil \log_2(\delta/\delta_0) \rceil. \tag{112}$$

The latter result proves the finiteness of required number of steps for finding the minimal α_0—thanks to the upper bound $\eta_1 + \eta_2$ on the required steps.

ACKNOWLEDGMENT

We would like to thank Prof. Antonio De Maio for providing us with the MATLAB code for SDR. Moreover, we would like to thank the associate editor and the four anonymous reviewers for their constructive and to-the-point comments that helped us significantly in clarifying the ideas of this work.

REFERENCES

[1] A. De Maio, S. De Nicola, Y. Huang, S. Zhang, and A. Farina, "Code design to optimize radar detection performance under accuracy and similarity constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5618–5629, Nov. 2008.

[2] A. De Maio and A. Farina, "Code selection for radar performance optimization," in *Proc. Waveform Divers. Design Conf.*, Pisa, Italy, Jun. 2007, pp. 219–223.

[3] H. He, J. Li, and P. Stoica, *Waveform Design for Active Sensing Systems: A Computational Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[4] N. Levanon and E. Mozeson, *Radar Signals*. New York, NY, USA: Wiley, 2004.

[5] H. He, P. Stoica, and J. Li, "On synthesizing cross ambiguity functions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3536–3539.

[6] , J. Li and P. Stoica, Eds., *Robust Adaptive Beamforming*. Hoboken, NJ, USA: Wiley, 2005.

[7] K.-C. Tan, G.-L. Oh, and M. Er, "A study of the uniqueness of steering vectors in array processing," *Signal Process.*, vol. 34, no. 3, pp. 245–256, 1993.

[8] A. Khabbazibasmenj, S. Vorobyov, and A. Hassanien, "Robust adaptive beamforming via estimating steering vector based on semidefinite relaxation," in *Proc. Conf. Signals, Syst. Comput. (ASILOMAR)*, CA, Nov. 2010, pp. 1102–1106.

[9] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Math. Program.*, vol. 142, no. 1–2, pp. 397–434, 2013.

[10] J. Jaldén, C. Martin, and B. Ottersten, "Semidefinite programming for detection in linear systems—Optimality conditions and space-time decoding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Hong Kong, Apr. 2003, vol. 4, pp. 9–12.

[11] S. Zhang and Y. Huang, "Complex quadratic optimization and semidefinite programming," *SIAM J. Optim.*, vol. 16, no. 3, pp. 871–890, 2006.

[12] A. T. Kyrillidis and G. N. Karystinos, "Rank-deficient quadratic-form maximization over M-phase alphabet: Polynomial-complexity solvability and algorithmic developments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2011, pp. 3856–3859.

[13] Z. Wen, D. Goldfarb, and K. Scheinberg, "Block coordinate descent methods for semidefinite programming," in *Handbook on Semidefinite, Conic and Polynomial Optimization*, ser. Int. Ser. Operat. Res. Manag. Sci., M. F. Anjos and J. B. Lasserre, Eds. New York, NY, USA: Springer, 2012, vol. 166, pp. 533–564.

[14] S. Verdú, "Computational complexity of optimum multiuser detection," *Algorithmica*, vol. 4, pp. 303–312, 1989.

[15] W.-K. Ma, B.-N. Vo, T. Davidson, and P.-C. Ching, "Blind ML detection of orthogonal space-time block codes: Efficient high-performance implementations," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 738–751, Feb. 2006.

[16] T. Cui and C. Tellambura, "Joint channel estimation and data detection for OFDM systems via sphere decoding," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, TX, Dec. 2004, vol. 6, pp. 3656–3660.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[18] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *ACM J.*, vol. 42, no. 6, pp. 1115–1145, Nov. 1995.

[19] A. De Maio, Y. Huang, M. Piezzo, S. Zhang, and A. Farina, "Design of optimized radar codes with a peak to average power ratio constraint," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2683–2697, Jun. 2011.

[20] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, 2010.

[21] A. So, J. Zhang, and Y. Ye, "On approximating complex quadratic optimization problems via semidefinite programming relaxations," *Math. Program.*, vol. 110, pp. 93–110, 2007.

[22] M. Kisialiou and Z.-Q. Luo, "Performance analysis of quasi-maximum-likelihood detector based on semi-definite programming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, PA, 2005, vol. 3, pp. iii–433.

[23] M. Soltanalian, B. Tang, J. Li, and P. Stoica, "Joint design of the receive filter and transmit sequence for active sensing," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 423–426, 2013.

[24] J. Tropp, I. Dhillon, R. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005.

[25] W. Glunt, T. L. Hayden, and R. Reams, "The nearest doubly stochastic matrix to a real matrix with the same first moment," *Numer. Linear Algebra With Appl.*, vol. 5, no. 6, pp. 475–482, 1998.

[26] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[27] G. Karystinos and A. Liavas, "Efficient computation of the binary vector that maximizes a rank-deficient quadratic form," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3581–3593, Jul. 2010.

[28] A. De Maio, Y. Huang, and M. Piezzo, "A Doppler robust max-min approach to radar code design," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4943–4947, Sep. 2010.

[29] P. Stoica, H. He, and J. Li, "New algorithms for designing unimodular sequences with good correlation properties," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1415–1425, Apr. 2009.

[30] M. Soltanalian and P. Stoica, "Computational design of sequences with good correlation properties," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2180–2193, May 2012.

[31] J. Benedetto, I. Konstantinidis, and M. Rangaswamy, "Phase-coded waveforms and their design," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 22–31, Jan. 2009.

Mojtaba Soltanalian (S'08) received the B.Sc. degree in electrical engineering (communications) from Sharif University of Technology, Tehran, Iran, in 2009.

He is currently working toward the Ph.D. degree in electrical engineering with applications in signal processing at the Department of Information Technology, Uppsala University, Sweden. His research interests include different aspects of sequence design for active sensing, communications, and biology.

Petre Stoica (SM'91–F'94) is currently a Professor of systems modeling at the Department of Information Technology, Uppsala University, Sweden. More details about him can be found at http://www.it.uu.se/katalog/ps.