Unified Optimization Framework for Multi-Static Radar Code Design using Information-Theoretic Criteria

Mohammad Mahdi Naghsh*, Mahmoud Modarres-Hashemi, Shahram ShahbazPanahi, *Senior Member, IEEE*, Mojtaba Soltanalian, and Petre Stoica, *Fellow, IEEE*

Abstract

In this paper, we study the problem of code design to improve the detection performance of multi-static radar in the presence of clutter (i.e., a signal-dependent interference). To this end, we briefly present a discrete-time formulation of the problem as well as the optimal detector in the presence of Gaussian clutter. Due to the lack of analytical expression for receiver operation characteristic (ROC), code design based on ROC is not feasible. Therefore, we consider several popular information-theoretic criteria including Bhattacharyya distance, Kullback-Leibler (KL) divergence, J-divergence, and mutual information (MI) as design metrics. The code optimization problems associated with different information-theoretic criteria are obtained and cast under a unified framework. We propose two general methods based on Majorization-Minimization to tackle the optimization problems in the framework. The first method provides optimal solutions via successive majorizations whereas the second one consists of a majorization step, a relaxation, and a synthesis stage. Moreover, derivations of the proposed methods are extended to tackle the code design problems with a peak-to-average ratio power (PAR) constraint. Using numerical investigations, a general analysis of the coded system performance, computational efficiency of the proposed methods, and the behavior of the information-theoretic criteria is provided.

This work was supported in part by the European Research Council (ERC) under Grant #228044 and the Swedish Research Council. M. M. Naghsh and M. Modarres-Hashemi are with the Dept. of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran. S. ShahbazPanahi is with the Faculty of Engineering and Applied Science of the University of Ontario Institute of Technology, 2000 Simcoe Street North, Oshawa, ON, Canada, L1H 7K4. M. Soltanalian and P. Stoica are with the Dept. of Information Technology, Uppsala University, Uppsala, SE 75105, Sweden.

* Please address all the correspondence to M. M. Naghsh, Phone: (+98) 311-3912450; Fax: (+98) 311-3912451; Email: mm_naghsh@ec.iut.ac.ir

Keywords: Code design, information-theoretic criteria, multi-static radar, majorization-minimization, peak-to-average power ratio (PAR)

I. Introduction

Signal design for detection performance improvement has been a long-term research topic in the radar literature. Active radars deal with both signal-dependent as well as signal-independent interferences. Indeed, the signals backscattered from undesired obstacles (known as clutter) depend on the transmit signal, whereas noise, unwanted emissions, and jammer emissions do not depend on the transmit signal.

The effect of the clutter has been considered in early studies for single-input single-output (SISO) systems [1]–[5]. The aim of these studies is to maximize the signal-to-interference-plus-noise-ratio (SINR) by means of joint optimization of the transmit signal and the receive filter. The results of [6] and [7] are recent extensions of [5] which use different methods to tackle some related problems. In [8], a solution for the optimal energy spectral density (ESD) of the transmit signal as well as a method for approximate synthesis of the signal have been presented for SISO systems. Problems related to that of [8] have been considered in [9] and [10] for cases where practical constraints such as low peak-to-average-power ratio (PAR) and similarity to a given code are imposed in the design stage. The work of [11] employs mutual information (MI) as design metric for target detection and estimation. In [12], two signal design approaches based on MI and SINR have been studied for extended target recognition in SISO systems. KL-divergence has been considered in [13] for target classification.

In multi-static scenarios, the interpretation of the detection performance is not easy in general and in several cases expressions for detection performance are too complicated to be amenable to utilization as design metrics (see e.g. [14] [15]). In such circumstances, information-theoretic criteria can be considered as design metrics to guarantee some types of optimality for the obtained signals. For example, in [15] an approach similar to that of [8] has been applied to the case of multi-static radars with one transmit antenna, and a concave approximation of the J-divergence has been used as the design metric. MI has been considered as a design metric for non-orthogonal multiple-input multiple-output (MIMO) radar signal design in [16] for clutter-free scenarios. A problem related to that of [16] has been studied in [18] where Kullback-Leibler (KL) divergence and J-divergence are used as design metrics. In [19], KL-divergence and MI have been taken into account for MIMO radar signal design in the absence of clutter. Information-theoretic criteria have also been used in research subjects related to the detection problem. The authors in [20] study the target classification for MIMO radars using minimum mean-square error (MMSE) and the MI criterion assuming no clutter. The reference [21] employs Bhatacharyya distance, KL-divergence, and J-divergence for signal design of a communication system with multiple transmit antennas. MI has also been used to investigate the effect of the jammer on MIMO radar performance in clutter-free situations in [22].

In this paper, we provide a unified framework for multi-static radar code design in the presence of clutter. Although closed-form expressions for the probability of detection and the probability of false alarm of the optimal detector are available, the analytical receiver operating characteristic (ROC) does not exist. As such, we employ several information-theoretic criteria that are widely used in the literature (see e.g. [16], [19], [21]), namely Bhattacharyya distance, KL-divergence, J-divergence, and MI as metrics for code design. In particular, we express these metrics in terms of the code vector and then present corresponding optimization problems. We show that the arising optimization problems can be conveniently dealt with using a unified framework. To tackle the code design problem, two novel methods based on Majorization-Minimization (MaMi) technique are devised. In the first method (which we call Sv-MaMi) successive majorizations are employed, whereas the second one (which we call Re-MaMi) is based on majorizations, a relaxation, and a synthesis stage. We also extend the proposed methods to the code design problem with PAR constraints and to the case of multiple transmitters (with orthogonal transmission). To the best of our knowledge, no study of code design with PAR constraints using information-theoretic criteria was conducted prior to this work.

The rest of this paper is organized as follows. In Section II, we present a discrete-time formulation of the detection problem as well as the optimal detector. We briefly review different information-theoretic criteria in Section III and cast the associated optimization problems under a unified framework. Section IV contains the derivations of the steps of Sv-MaMi to deal with the optimization problems formulated in the unified framework presented in Section III. Re-MaMi is introduced in Section V as an alternative approach to obtain optimized codes of the arising optimization problems. Extensions of the design problem to the cases of PAR-constrained design and multiple transmitters (with orthogonal transmission) are discussed in Section VI. Numerical examples are provided in Section VII. Finally, conclusions are drawn in Section VIII.

Notation: We use bold lowercase letters for vectors and bold uppercase letters for matrices. $(.)^T$, $(.)^*$ and $(.)^H$ denote the vector/matrix transpose, the complex conjugate, and the Hermitian transpose, respectively. $\mathbf{I}_N$ represents the identity matrix in $\mathbb{C}^{N\times N}$. $\|\mathbf{X}\|_F$ denotes the Frobenius norm of a matrix $\mathbf{X}$. The notations $\mu_{max}(\cdot)$ and $\mu_{min}(\cdot)$ indicate the principal and the minor eigenvalues of a Hermitian matrix, respectively. The l_2-norm of a vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\|_2$. tr$(.)$ is the trace of a square matrix argument. $\mathrm{blkDiag}(.)$ denotes the block diagonal matrix formed by its arguments. We write $\mathbf{A} \succeq \mathbf{B}$ iff

$\mathbf{A} - \mathbf{B}$ is positive semi-definite, and $\mathbf{A} \succ \mathbf{B}$ iff $\mathbf{A} - \mathbf{B}$ is positive-definite. $\mathrm{E}\{.\}$ stands for the statistical expectation operator. $\mathcal{CN}(\omega, \mathbf{\Sigma})$ denotes the circularly symmetric complex Gaussian distribution with mean ω and covariance $\mathbf{\Sigma}$. The symbol $\sim$ is used to show the distribution of a random variable/vector. Finally, $\Re(.)$ denotes the real-part of the complex-valued argument.

II. Data Modeling and the Optimal Detector

A. Data Modeling

We consider a multi-static pulsed-radar with one transmitter and N_r widely separated receive antennas. The baseband transmit signal can be formulated as

$$s(t) = \sum_{n=1}^{N} a_n \phi(t - [n-1]T_P) \tag{1}$$

where $\phi(t)$ is the basic unit-energy transmit pulse (with time duration τ_p), T_P is the pulse repetition period ($T_P \gg \tau_p$), and $\{a_n\}_{n=1}^{N}$ are the deterministic coefficients that are to be "optimally" determined. The vector $\mathbf{a} \triangleq [a_1 \;\; a_2 \;\; \ldots \;\; a_N]^T$ is referred to as the code vector of the radar system.

The baseband signal received at the k^{th} antenna backscattered from a stationary target can be written as

$$r_k(t) = \alpha_k s(t - \tau_k) + c_k(t) + w_k(t) \tag{2}$$

where α_k is the amplitude of the target return (including the channel effects), $c_k(t)$ is the clutter component, $w_k(t)$ is a Gaussian random process representing the signal-independent interference component (including various types of noise, interference, and jamming), and τ_k is the time corresponding to propagation delay for the path from the transmitter to the target and thereafter to the k^{th} receiver.

In what follows, we consider a few typical assumptions in the radar literature which are key to the derivations that will appear in this paper.

Assumption 1. *We assume that the clutter component at the k^{th} receiver is composed of signal echoes produced by many stationary point scatterers (located within unambiguous-range with respect to the k^{th} receiver [24]). The amplitudes and arrival times of the echoes are assumed to be statistically independent [25], [26].*

According to Assumption 1, the clutter component can be expressed as

$$c_k(t) = \sum_{v=1}^{N_c} \rho_{k,v} s(t - \tau_{k,v}) \tag{3}$$

where N_c is the number of point scatterers, $\rho_{k,v}$ is the "amplitude" of the v^{th} scatterer observed by the k^{th} receive antenna, and $\tau_{k,v}$ is the propagation delay at the k^{th} receiver corresponding to the v^{th} scatterer for which we have $\tau_{k,v} \leq T_p$.

At the k^{th} receiver, the received signal is matched filtered by $\phi^*(-t)$. Then range-gating is performed by sampling the output of the matched filter at time slots corresponding to a specific radar cell. Note that the detection for a specific radar cell can be accomplished using a successive chain of operations including directional transmission and reception as well as range-gating at each receiver [27].

The discrete-time signal corresponding to a certain radar cell for the k^{th} receiver can be described as (see Appendix A):

$$r_{k,n} = \alpha_k a_n + \widetilde{\rho}_k a_n + w_{k,n} \tag{4}$$

where $r_{k,n}$ is the output of the matched filter at the k^{th} receiver sampled at $t = (n-1)T_p + \tau_k$, $\widetilde{\rho}_k$ is a zero-mean complex Gaussian random variable (RV) with variance $\sigma^2_{\mathrm{c},k}$ associated with the clutter scatterers, and $w_{k,n}$ denotes the n^{th} sample of $w_k(t)$ when filtered by $\phi^*(-t)$ at the k^{th} receiver. Using a vector notation, we can write

$$\mathbf{r}_k \triangleq \mathbf{s}_k + \mathbf{c}_k + \mathbf{w}_k = \alpha_k \mathbf{a} + \widetilde{\rho}_k \mathbf{a} + \mathbf{w}_k \tag{5}$$

where $\mathbf{r}_k \triangleq [r_{k,1} \;\; r_{k,2} \;\; \cdots \;\; r_{k,N}]^T$, $\mathbf{w}_k \triangleq [w_{k,1} \;\; w_{k,2} \;\; \cdots \;\; w_{k,N}]^T$, $\mathbf{s}_k \triangleq \alpha_k \mathbf{a}$, and $\mathbf{c}_k \triangleq \widetilde{\rho}_k \mathbf{a}$.

We further make the following assumptions:

Assumption 2. *The Swerling-I model is used for the amplitude of the target echo, i.e.* $\alpha_k \sim \mathcal{CN}(0, \sigma^2_k)$ *for any stationary target [8], [15].*

Assumption 3. *The second-order statistics of the target, clutter, and interference components at the* k^{th} *receiver (i.e.* $\sigma^2_k, \sigma^2_{\mathrm{c},k}$, *and* $\mathrm{E}\{\mathbf{w}_k \mathbf{w}_k^H\}$*) are assumed to be known.*

The above assumption is common for radar systems using cognitive (knowledged-aided) methods that employ geographical, meteorological, National Land Cover Data (NLCD), and the information of the previous scan to interactively learn and extract the characteristics of the environment (see e.g. [10], [28]–[31]).

Assumption 4. *The random variables in the set* $\{\alpha_k\}_{k=1}^{N_r}$ *are statistically independent. Such a statistical independence is also considered for random variables/vectors in the sets* $\{\widetilde{\rho}_k\}_{k=1}^{N_r}$ *and* $\{\mathbf{w}_k\}_{k=1}^{N_r}$.

Assumption 4 is well-justified, due to the fact that the receivers are widely separated [15], [16].

B. Optimal Detector

Using all the received signals, the target detection leads to the following binary hypothesis problem

$$\begin{cases} \mathrm{H}_0: & \mathbf{r} = \mathbf{c} + \mathbf{w} \\ \mathrm{H}_1: & \mathbf{r} = \mathbf{s} + \mathbf{c} + \mathbf{w} \end{cases} \tag{6}$$

where $\mathbf{r}$, $\mathbf{s}$, $\mathbf{c}$, and $\mathbf{w}$ are defined by column-wise stacking of $\mathbf{r}_k, \mathbf{s}_k, \mathbf{c}_k$, and $\mathbf{w}_k$ for $k = 1, 2, ..., N_r$; more precisely, $\mathbf{r} \triangleq [\mathbf{r}_1^T \ \cdots \ \mathbf{r}_{N_r}^T]^T$, $\mathbf{s} \triangleq [\mathbf{s}_1^T \ \cdots \ \mathbf{s}_{N_r}^T]^T$, $\mathbf{c} \triangleq [\mathbf{c}_1^T \ \cdots \ \mathbf{c}_{N_r}^T]^T$, and $\mathbf{w} \triangleq [\mathbf{w}_1^T \ \cdots \ \mathbf{w}_{N_r}^T]^T$.

Let $\{\mathbf{M}_k\}$ denote the covariance matrices of Gaussian random vectors $\{\mathbf{w}_k\}$. Further let $\mathbf{S}, \mathbf{C}$, and $\mathbf{M}$ represent the the covariance matrices of $\mathbf{s}, \mathbf{c}$, and $\mathbf{w}$, respectively. Using the aforementioned assumptions we have that

$$\begin{aligned} \mathbf{S} &= \mathrm{blkDiag}(\sigma_1^2 \mathbf{a}\mathbf{a}^H, \sigma_2^2 \mathbf{a}\mathbf{a}^H, \ldots, \sigma_{N_r}^2 \mathbf{a}\mathbf{a}^H) \\ \mathbf{C} &= \mathrm{blkDiag}(\sigma_{\mathrm{c},1}^2 \mathbf{a}\mathbf{a}^H, \sigma_{\mathrm{c},2}^2 \mathbf{a}\mathbf{a}^H, \ldots, \sigma_{\mathrm{c},N_r}^2 \mathbf{a}\mathbf{a}^H) \\ \mathbf{M} &= \mathrm{blkDiag}(\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_{N_r}). \end{aligned} \tag{7}$$

Consequently, the underlying detection problem can be equivalently expressed as

$$\begin{cases} \mathrm{H}_0: \mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}) \\ \mathrm{H}_1: \mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{DSD} + \mathbf{I}) \end{cases} \tag{8}$$

where $\mathbf{x} \triangleq \mathbf{Dr}$, $\mathbf{D} \triangleq (\mathbf{C} + \mathbf{M})^{-\frac{1}{2}} = \mathrm{blkDiag}(\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{N_r})$ with $\mathbf{D}_k \triangleq (\sigma_{\mathrm{c},k}^2 \mathbf{a}\mathbf{a}^H + \mathbf{M}_k)^{-\frac{1}{2}}$. Note that both $\mathbf{D}$ and $\mathbf{S}$ in (8) depend on the transmit code $\mathbf{a}$. The optimal detector for (8) can be obtained by applying the estimator-correlator theorem [32, chapter 13] as:

$$\sum_{k=1}^{N_r} \sigma_k^2 \mathbf{x}_k^H \mathbf{D}_k \mathbf{a}\mathbf{a}^H \mathbf{D}_k (\sigma_k^2 \mathbf{D}_k \mathbf{a}\mathbf{a}^H \mathbf{D}_k + \mathbf{I})^{-1} \mathbf{x}_k \underset{\mathrm{H}_1}{\overset{\mathrm{H}_0}{\lessgtr}} \eta \tag{9}$$

where η is the detection threshold, and

$$\mathbf{x}_k = \mathbf{D}_k \mathbf{r}_k \,. \tag{10}$$

In particular, by defining

$$\begin{aligned} \lambda_k &\triangleq \sigma_k^2 \mathbf{a}^H \mathbf{D}_k^2 \mathbf{a}, \\ \theta_k &\triangleq \frac{\mathbf{a}^H \mathbf{D}_k \mathbf{x}_k}{\|\mathbf{a}^H \mathbf{D}_k\|_2}, \end{aligned} \tag{11}$$

the canonical form of the detector in (9) can be obtained as

$$T(\boldsymbol{\theta}) \triangleq \sum_{k=1}^{N_r} \frac{\lambda_k |\theta_k|^2}{1 + \lambda_k} \underset{\mathrm{H}_1}{\overset{\mathrm{H}_0}{\lessgtr}} \eta \tag{12}$$

where $\boldsymbol{\theta} \triangleq [\theta_1 \ \ \theta_2 \ \ \cdots \ \ \theta_{N_r}]^T$.

III. Optimal Code Design

In this section, we aim to obtain the optimal transmit signals by judiciously designing the code vector $\mathbf{a}$. A reasonable approach to code design is to exploit the knowledge of the analytical receiver operating characteristic (ROC) which enables the designer to obtain the largest possible value of the probability of detection P_d for a given value of the probability of false alarm P_{fa} via optimal selection of the design parameters. However, this method cannot be used if the analytical ROC is not amenable to a closed-form expression which is the case for the problem considered in this paper. Particularly, even though closed-form expressions for P_d and P_{fa} can be obtained by applying the results of [14], derivation of the analytical ROC is not possible. In such cases, one can resort to information-theoretic criteria including Bhattacharyya distance, KL-divergence, J-divergence, and MI (see the Introduction). In what follows, the goal is to improve the detection performance by maximizing the aforementioned information-theoretic criteria over the code vector $\mathbf{a}$. Interestingly, the corresponding optimization problems can be dealt with conveniently using a unified optimization framework.

A. Information-theoretic design metrics

• **Bhattacharyya distance**: Bhattacharyya distance $\mathcal{B}$ measures the distance between two probability density functions (pdf). In a binary hypothesis testing problem $T(H_0, H_1)$, the design parameters can be chosen such that the Bhattacharyya distance $\mathcal{B}$ between the pdfs of the observation under H_0 and H_1 is maximized. Note that the Bhattacharyya distance provides an upper bound on P_{fa}, and at the same time yields a lower bound on P_d [21]. Therefore, maximization of the Bhattacharyya distance minimizes the upper bound on P_{fa} while it maximizes the lower bound on P_d[1].

The Bhattacharyya distance $\mathcal{B}$ for two multivariate Gaussian distributions, $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_1)$ and $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_2)$, can be expressed as [21]:

$$\mathcal{B} = \log\left(\frac{\det(0.5(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2))}{\sqrt{\det(\mathbf{\Sigma}_1)\det(\mathbf{\Sigma}_2)}}\right). \tag{13}$$

By applying (13) to the problem in (8) we obtain

$$\begin{aligned}\mathcal{B} &= \log\left(\frac{\det(\mathbf{I} + 0.5\mathbf{DSD})}{\sqrt{\det(\mathbf{I} + \mathbf{DSD})}}\right) \\ &= \sum_{k=1}^{N_r} \log\left(\frac{\det(\mathbf{I} + 0.5\sigma_k^2 \mathbf{D}_k \mathbf{a}\mathbf{a}^H \mathbf{D}_k)}{\sqrt{\det(\mathbf{I} + \sigma_k^2 \mathbf{D}_k \mathbf{a}\mathbf{a}^H \mathbf{D}_k)}}\right) = \sum_{k=1}^{N_r} \log\left(\frac{1 + 0.5\lambda_k}{\sqrt{1+\lambda_k}}\right).\end{aligned} \tag{14}$$

[1]This is due to the fact that $P_d \geq 1 - \delta^{-1/2} e^{-\mathcal{B}}$ and $P_{fa} \leq \delta^{1/2} e^{-\mathcal{B}}$ where δ is the likelihood threshold [33].

The second equality in (14) holds due to the block-diagonal structure of the matrices $\mathbf{S}$ and $\mathbf{D}$. The last equality follows from the fact that the eigenvalues of the matrix $\mathbf{I}+\sigma_k^2\mathbf{D}_k\mathbf{a}\mathbf{a}^H\mathbf{D}_k$ include $(N-1)$ ones and the maximum eigenvalue which is given by $1+\sigma_k^2\mathbf{a}^H\mathbf{D}_k^2\mathbf{a} = 1+\lambda_k$. Eventually the underlying code design problem can be formulated as

$$\begin{aligned} \max_{\mathbf{a},\lambda_k} \quad & \sum_{k=1}^{N_r} \log\left(\frac{1+0.5\lambda_k}{\sqrt{1+\lambda_k}}\right) \\ \text{subject to} \quad & \lambda_k = \sigma_k^2\mathbf{a}^H(\sigma_{\mathrm{c},k}^2\mathbf{a}\mathbf{a}^H+\mathbf{M}_k)^{-1}\mathbf{a} \\ & \|\mathbf{a}\|_2^2 \leqslant e, \end{aligned} \tag{15}$$

where e denotes the total transmit energy.

- **KL-divergence**: The KL-divergence $\mathcal{D}(f_0\|f_1)$ is another metric to measure the "distance" between two pdfs f_0 and f_1. Consider a binary hypothesis testing problem with $f_0 = f(\mathbf{r}|\mathrm{H}_0)$ and $f_1 = f(\mathbf{r}|\mathrm{H}_1)$. The Stein Lemma states that for any fixed P_{fa} [21]

$$\mathcal{D}\left(f(\mathbf{r}|\mathrm{H}_0)\|f(\mathbf{r}|\mathrm{H}_1)\right) = \lim_{N\to\infty}\left(-\frac{1}{N}\log(1-P_d)\right) \tag{16}$$

which implies that (for any fixed P_{fa}) the maximization of the KL-divergence metric leads to an asymptotic maximization of P_d (we refer the interested reader to [34, Chapter 4], [21], [35, Theorem 1], and references therein for some bounds on the detection performance associated with the KL-divergence). In addition, we have that [21]

$$\mathcal{D}(f(\mathbf{r}|\mathrm{H}_0)\|f(\mathbf{r}|\mathrm{H}_1)) = -\,\mathrm{E}\left\{\log\left(\mathcal{L}\right)|\mathrm{H}_0\right\} \tag{17}$$

where $\mathcal{L}$ is the likelihood ratio defined as

$$\mathcal{L} \triangleq \frac{f(\mathbf{r}|\mathrm{H}_1)}{f(\mathbf{r}|\mathrm{H}_0)}\,.$$

Using (12), (17) and the identity $\log\left(\mathcal{L}\right) = T(\boldsymbol{\theta}) - \log\det(\mathbf{I}+\mathbf{DSD})$ [32], the KL-divergence associated with (8) can be obtained as

$$\mathcal{D}\left(f(\mathbf{r}|\mathrm{H}_0)\|f(\mathbf{r}|\mathrm{H}_1)\right) = \sum_{k=1}^{N_r}\left\{\log(1+\lambda_k) - \frac{\lambda_k}{1+\lambda_k}\right\}\,.$$

As a result, the problem of code design by maximizing the KL-divergence metric can be stated as:

$$\begin{aligned} \max_{\mathbf{a},\lambda_k} \quad & \sum_{k=1}^{N_r}\left\{\log(1+\lambda_k) - \frac{\lambda_k}{1+\lambda_k}\right\} \\ \text{subject to} \quad & \lambda_k = \sigma_k^2\mathbf{a}^H(\sigma_{\mathrm{c},k}^2\mathbf{a}\mathbf{a}^H+\mathbf{M}_k)^{-1}\mathbf{a} \\ & \|\mathbf{a}\|_2^2 \leqslant e. \end{aligned} \tag{18}$$

• **J-divergence**: The J-divergence metric, denoted herein as $\mathcal{J}$, is another measure of the distance between two pdfs and it is defined as

$$\mathcal{J} \triangleq \mathcal{D}(f_0\|f_1) + \mathcal{D}(f_1\|f_0). \tag{19}$$

According to Stein Lemma [34, Chapter 4], in a binary hypothesis testing problem (with $f_0 = f(\mathbf{r}|\mathrm{H}_0)$ and $f_1 = f(\mathbf{r}|\mathrm{H}_1)$), and for any fixed P_d, we can write

$$\mathcal{D}\left(f(\mathbf{r}|\mathrm{H}_1)\|f(\mathbf{r}|\mathrm{H}_0)\right) = \lim_{N\to\infty}\left(-\frac{1}{N}\log(P_{fa})\right). \tag{20}$$

Eq. (16) and (20) along with other properties and bounds associated with the J-divergence (see e.g [34, Chapter 4], [35, Theorem 1], [36], [21], and references therein) have motivated several authors to consider $\mathcal{J}$ as the design metric for radar signal design (see [15], [21], [37], and references therein). For the binary hypothesis testing problem in (8) with $f_0 = f(\mathbf{r}|\mathrm{H}_0)$ and $f_1 = f(\mathbf{r}|\mathrm{H}_1)$, we have that [21]

$$\begin{aligned}\mathcal{J} &= \mathrm{E}\{(\mathcal{L}-1)\log(\mathcal{L})|\mathrm{H}_0\}\\ &= \left(\int \frac{f(\mathbf{r}|\mathrm{H}_1)}{f(\mathbf{r}|\mathrm{H}_0)}\log(\mathcal{L})f(\mathbf{r}|\mathrm{H}_0)d\mathbf{r}\right) - \mathrm{E}\{\log(\mathcal{L})|H_0\}\\ &= \mathrm{E}\{\log(\mathcal{L})|H_1\} - \mathrm{E}\{\log(\mathcal{L})|H_0\}.\end{aligned}$$

Using (21) along with similar calculations as in the case of KL-divergence, the J-divergence metric associated with (8) can be obtained as

$$\mathcal{J} = \sum_{k=1}^{N_r}\frac{\lambda_k^2}{1+\lambda_k}. \tag{21}$$

Consequently, the corresponding code design problem can be expressed as

$$\begin{aligned}\max_{\mathbf{a},\lambda_k} \quad & \sum_{k=1}^{N_r}\frac{\lambda_k^2}{1+\lambda_k}\\ \text{subject to} \quad & \lambda_k = \sigma_k^2\mathbf{a}^H(\sigma_{\mathrm{c},k}^2\mathbf{a}\mathbf{a}^H + \mathbf{M}_k)^{-1}\mathbf{a}\\ & \|\mathbf{a}\|_2^2 \leqslant e.\end{aligned} \tag{22}$$

• **Mutual information**: MI is another metric that has been used for radar transmit signal design (see the Introduction). The MI between the amplitude of the target return and the received signal is often considered as a design criterion. For the relationship between MI and minimum mean-square error (MMSE) estimation see e.g. [38]. Note that the larger the MI the better the MMSE estimation [39, Chapter 2]. Note also that, the optimal detector for Gaussian pdfs has a close relationship to the MMSE estimation (see e.g. the estimator-correlator theorem in [32, Chapter 5 and 13]) in the sense that better estimation leads to detection performance improvements [16]. Furthermore, a comprehensive mathematical motivation for using MI as a metric in radar signal design is provided in [39, Chapter 2] and [11] using rate-distortion function, Fano's inequality, and Shannon's noisy channel coding theorem. Additionally, the results of [40] relate the MI and Bayes risk in statistical decision problems. An analysis of the connection between Bayesian classification performance and MI has also been performed in [41]. The MI metric associated with (8) is given by [42]

$$\begin{aligned}\mathcal{M} &= \log\left((\pi e)^N \det(\mathbf{I}+\mathbf{DSD})\right) - \log\left((\pi e)^N \det(\mathbf{I})\right)\\ &= \sum_{k=1}^{N_r} \log\left(\det(\mathbf{I}+\sigma_k^2 \mathbf{D}_k \mathbf{a}\mathbf{a}^H \mathbf{D}_k)\right) - \sum_{k=1}^{N_r} \log\left(\det(\mathbf{I})\right)\\ &= \sum_{k=1}^{N_r} \log(1+\lambda_k) \end{aligned} \quad (23)$$

where the second equality follows from the block-diagonal structures of $\mathbf{D}$ and $\mathbf{S}$, and the third equality holds due to the fact that $\left\{\sigma_k^2 \mathbf{D}_k \mathbf{a}\mathbf{a}^H \mathbf{D}_k\right\}$ are rank-one. Therefore, the $\mathcal{M}$-optimal code $\mathbf{a}$ is the solution to the following maximization problem:

$$\begin{aligned} \max_{\mathbf{a},\lambda_k} \quad & \sum_{k=1}^{N_r} \log(1+\lambda_k)\\ \text{subject to} \quad & \lambda_k = \sigma_k^2 \mathbf{a}^H (\sigma_{\mathrm{c},k}^2 \mathbf{a}\mathbf{a}^H + \mathbf{M}_k)^{-1}\mathbf{a} \\ & \|\mathbf{a}\|_2^2 \leqslant e. \end{aligned} \quad (24)$$

B. *Unified Framework*

Herein we cast the optimization problems corresponding to various information-theoretic criteria discussed earlier under a unified optimization framework. Indeed, we consider the following general form of the optimization problems in (15), (18), (22), and (24):

$$\begin{aligned} \max_{\mathbf{a},\lambda_k} \quad & \sum_{k=1}^{N_r} f_{\mathcal{I}}(\lambda_k) + g_{\mathcal{I}}(\lambda_k)\\ \text{subject to} \quad & \lambda_k = \sigma_k^2 \mathbf{a}^H (\sigma_{\mathrm{c},k}^2 \mathbf{a}\mathbf{a}^H + \mathbf{M}_k)^{-1}\mathbf{a} \\ & \|\mathbf{a}\|_2^2 \leqslant e, \end{aligned} \quad (25)$$

where $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$, and $f_{\mathcal{I}}(.)$ and $g_{\mathcal{I}}(.)$ are concave and convex functions for any $\mathcal{I}$, respectively. More precisely, we have that

$$\begin{cases} f_{\mathcal{B}}(\lambda_k) = \log(1 + 0.5\lambda_k), & g_{\mathcal{B}}(\lambda_k) = -\frac{1}{2}\log(1+\lambda_k), \\ f_{\mathcal{D}}(\lambda_k) = \log(1+\lambda_k), & g_{\mathcal{D}}(\lambda_k) = \frac{1}{1+\lambda_k} - 1, \\ f_{\mathcal{J}}(\lambda_k) = 0, & g_{\mathcal{J}}(\lambda_k) = \frac{\lambda_k^2}{1+\lambda_k}, \\ f_{\mathcal{M}}(\lambda_k) = \log(1+\lambda_k), & g_{\mathcal{M}}(\lambda_k) = 0. \end{cases}$$

Remark 1: In the case of spatially wide-sense stationary (up to a power scale) signal-independent interferences, we have that $\mathbf{M}_k = \sigma_{w,k}^2 \widetilde{\mathbf{M}}$, $k = 1, 2, .., N_r$ (see e.g. [16]). In such a situation, a closed-form solution to the optimization problem (25) can be obtained. In particular, note that for any $\mathbf{M}_k$, a simplified expression of λ_k can be obtained using the matrix inversion lemma as:

$$\begin{aligned} \lambda_k &= \sigma_k^2 \mathbf{a}^H \left(\mathbf{M}_k^{-1} - \sigma_{\mathrm{c,k}}^2 \frac{\mathbf{M}_k^{-1}\mathbf{a}\mathbf{a}^H\mathbf{M}_k^{-1}}{1 + \sigma_{\mathrm{c,k}}^2 \mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}} \right) \mathbf{a} \\ &= \frac{\sigma_k^2 \mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}}{1 + \sigma_{\mathrm{c,k}}^2 \mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}}. \end{aligned} \tag{26}$$

Therefore, one can recast (25) as the following optimization problem:

$$\begin{aligned} \max_{\mathbf{a}} \quad & \sum_{k=1}^{N_r} q_{\mathcal{I}} \left(\frac{\sigma_k^2 \mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}}{1 + \sigma_{\mathrm{c,k}}^2 \mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}} \right) \\ \text{subject to} \quad & \|\mathbf{a}\|_2^2 \le e, \end{aligned} \tag{27}$$

where $q_{\mathcal{I}}(.) = f_{\mathcal{I}}(.) + g_{\mathcal{I}}(.)$. Let $v_\star$ denote the optimal value of (27). We have that

$$v_\star \le \sum_{k=1}^{N_r} q_{\mathcal{I}}(\widetilde{\mathbf{a}}_{k,\mathcal{I}}) \tag{28}$$

with $\widetilde{\mathbf{a}}_{k,\mathcal{I}}$ being the maximizer of $q_{\mathcal{I}}(.)$ subject to $\|\mathbf{a}\|^2 \le e$ for fixed k. Now we claim that $\mathbf{a}_\star = \sqrt{e}\mathbf{u}$ with $\mathbf{u}$ being the minor eigenvector of $\widetilde{\mathbf{M}}$ is an optimal solution to the optimization problem in (27). To observe this fact, note that the $\mathbf{a}_\star$ maximizes λ_k, for any k, subject to the energy constraint because λ_k is an increasing function of $\mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}$. Moreover, $q_{\mathcal{I}}(\lambda_k)$ is an increasing function of λ_k for all $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$ (see Appendix B) and hence $\widetilde{\mathbf{a}}_{k,\mathcal{I}} = \mathbf{a}_\star$ for all $k, \mathcal{I}$. Consequently, $\mathbf{a} = \mathbf{a}_\star$ yields the upper bound on $v_\star$ for all $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$. ■

We use the Majorization-Minimization (or Minorization-Maximization) techniques to tackle the non-convex problems in (25). Majorization-Minimization (MaMi) is an iterative technique that can be used for obtaining a solution to the general minimization problem [43] [44]:

$$\begin{aligned} \min_{\mathbf{z}} \quad & \widetilde{f}(\mathbf{z}) \\ \text{subject to} \quad & c(\mathbf{z}) \leq 0 \end{aligned} \tag{29}$$

where $\widetilde{f}(.)$ and $c(.)$ are non-convex functions. Each iteration (say the l^{th} iteration) of MaMi consists of two steps:

- Majorization Step: Finding $p^{(l)}(\mathbf{z})$ such that its minimization is simpler than that of $\widetilde{f}(\mathbf{z})$ and $p^{(l)}(\mathbf{z})$ majorizes $\widetilde{f}(\mathbf{z})$, i.e.,

$$\begin{aligned} p^{(l)}(\mathbf{z}) &\geq \widetilde{f}(\mathbf{z}), \ \ \forall \mathbf{z} \\ p^{(l)}\left(\mathbf{z}^{(l-1)}\right) &= \widetilde{f}\left(\mathbf{z}^{(l-1)}\right) \end{aligned} \tag{30}$$

  with $\mathbf{z}^{(l-1)}$ being the value of $\mathbf{z}$ at the $(l-1)^{th}$ iteration.
- Minimization Step: Solving the optimization problem,

$$\begin{aligned} \min_{\mathbf{z}} \quad & p^{(l)}(\mathbf{z}) \\ \text{subject to} \quad & c(\mathbf{z}) \leq 0. \end{aligned} \tag{31}$$

  to obtain $\mathbf{z}^{(l)}$.

Note that applying the Majorization-Minimization technique to the optimization problem in (29) decreases the value of the objective function $\widetilde{f}(\mathbf{z})$ at each iteration. Indeed, we have

$$\begin{aligned} \widetilde{f}\left(\mathbf{z}^{(l-1)}\right) = p^{(l)}\left(\mathbf{z}^{(l-1)}\right) &\geq p^{(l)}\left(\mathbf{z}^{(l)}\right) \\ &\geq \widetilde{f}\left(\mathbf{z}^{(l)}\right). \end{aligned} \tag{32}$$

The first inequality above follows from the minimization step in (31) and the second inequality holds true due to the majorization step in (30). The descent property in (32) guarantees the convergence of the sequence $\{\widetilde{f}(\mathbf{z}^{(l)})\}$ (assuming $\widetilde{f}(\mathbf{z})$ is bounded from below). Generally, the goodness of the obtained solution (i.e. $\mathbf{z}$ after the convergence) depends on the employed starting point. The optimality of the obtained solution $\mathbf{z}$ has been addressed in [43]–[45], where the solution $\mathbf{z}$ was shown to be a stationary point of $\widetilde{f}(\mathbf{z})$ (under some mild conditions). It is worth mentioning that a similar monotonically increasing behavior is guaranteed for Minorization-Maximization technique. Such a behavior of the values of the objective function is important when considering the objective as a measure of the code performance.

Remark 2: Note that the objective function $\sum_{k=1}^{N_r} q_{\mathcal{I}}(\lambda_k)$ in the problem (25) is bounded from above. To observe this fact, note that λ_k for all k can be upper bounded (considering (26)) as

$$\lambda_k = \frac{\sigma_k^2 \mathbf{a}^H \mathbf{M}_k^{-1} \mathbf{a}}{1+\sigma_{\mathrm{c,k}}^2 \mathbf{a}^H \mathbf{M}_k^{-1} \mathbf{a}} \leq \frac{\sigma_k^2 \|\mathbf{a}\|_2^2 \mu_{max}(\mathbf{M}_k^{-1})}{1+\sigma_{\mathrm{c,k}}^2 \|\mathbf{a}\|_2^2 \mu_{min}(\mathbf{M}_k^{-1})} \leq \frac{\sigma_k^2\, e\, \mu_{min}^{-1}(\mathbf{M}_k)}{1+\sigma_{\mathrm{c,k}}^2\, e\, \mu_{max}^{-1}(\mathbf{M}_k)}. \tag{33}$$

Due to the fact that $q_{\mathcal{I}}(\lambda_k)$ is a monotonically increasing function of λ_k for all k and $\mathcal{I}$ (see Appendix B), the above equation leads to an upper bound on the objective function in (25) for all k and $\mathcal{I}$. ■

In the following sections, we propose two novel algorithms based on MaMi to yield optimized solutions to (25).

IV. Optimal Code Design Using successive Majorizations

In this section, we propose a novel algorithm based on successive majorizations (which we call Sv-MaMi) to obtain an optimal code $\mathbf{a}$. In particular, we apply successive majorizations to the optimization problem in (25) and show the following:

Theorem 1. *(Sv-MaMi algorithm) The solution $\mathbf{a} = \mathbf{a}_\star$ of (25) can be obtained iteratively by solving the following convex quadratically constrained quadratic program (QCQP) (at the $(l+1)^{th}$ iteration):*

$$\begin{aligned} \min_{\mathbf{a}} \quad & \mathbf{a}^H \left(\sum_{k=1}^{N_r} \phi_{k,\mathcal{I}}^{(l)} \mathbf{M}_k^{-1} \right) \mathbf{a} - \Re \left(\left(\sum_{k=1}^{N_r} \mathbf{d}_{k,\mathcal{I}}^{(l)} \right)^H \mathbf{a} \right) \\ \textit{subject to} \quad & \|\mathbf{a}\|_2^2 \leqslant e, \end{aligned} \tag{34}$$

where the positive constant $\{\phi_{k,\mathcal{I}}^{(l)}\}$ and the vectors $\{\mathbf{d}_{k,\mathcal{I}}^{(l)}\}$ depend on $\mathcal{I} \in \{\mathcal{B},\mathcal{D},\mathcal{J},\mathcal{M}\}$ and will be given below.

In the sequel, we provide a proof of Theorem 1. We begin by noting that the convex function $g_{\mathcal{I}}(.)$ can be minorized using its supporting hyperplane at any given $\lambda_k = \lambda_k^{(l)}$, viz.

$$g_{\mathcal{I}}(\lambda_k) \geq g_{\mathcal{I}}\left(\lambda_k^{(l)}\right) + g'_{\mathcal{I}}\left(\lambda_k^{(l)}\right)\left(\lambda_k - \lambda_k^{(l)}\right) \tag{35}$$

which implies that

$$\sum_{k=1}^{N_r} g_{\mathcal{I}}(\lambda_k) \geq \sum_{k=1}^{N_r} g_{\mathcal{I}}\left(\lambda_k^{(l)}\right) + \sum_{k=1}^{N_r} g'_{\mathcal{I}}\left(\lambda_k^{(l)}\right)\left(\lambda_k - \lambda_k^{(l)}\right). \tag{36}$$

Herein $\lambda_k^{(l)}$ denote the λ_k obtained at the l^{th} iteration and $g'_{\mathcal{I}}(.)$ denote the first-order derivative of $g_{\mathcal{I}}(.)$ for $\mathcal{I} \in \{\mathcal{B},\mathcal{D},\mathcal{J},\mathcal{M}\}$.

Now observe that using (26), the optimal code $\mathbf{a} = \mathbf{a}_\star$ can be obtained in an iterative manner solving the following maximization at the $(l+1)^{th}$ iteration:

$$\max_{\mathbf{a},\lambda_k} \quad \sum_{k=1}^{N_r} f_{\mathcal{I}}(\lambda_k) + g'_{\mathcal{I}}\left(\lambda_k^{(l)}\right)\ \lambda_k \tag{37}$$

$$\text{subject to} \quad \lambda_k = \gamma_k - \frac{\gamma_k}{1+\beta_k \mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}} \tag{38}$$

$$\|\mathbf{a}\|_2^2 \leq e, \tag{39}$$

where $\gamma_k = \frac{\sigma_k^2}{\sigma_{\mathrm{c,k}}^2}$ and $\beta_k = \sigma_{\mathrm{c,k}}^2$. Note that the above problem is non-convex due to the non-affine equality constraint (38). The following Lemmas pave the way toward the derivation of the convex QCQPs of Theorem 1 corresponding to $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$.

Lemma 1. *If $f(x)$ is twice differentiable and if there exists $U > 0$ such that $f''(x) \leq U$ for all x, then for any given $\tilde{x}$, the convex quadratic function*

$$f(\tilde{x}) + f'(\tilde{x})(x-\tilde{x}) + \frac{1}{2}U(x-\tilde{x})^2 \tag{40}$$

majorizes f at $\tilde{x}$.

Proof: See [44, Section 3.4]. ■

Lemma 2. *Let $f(x)$ be an even function (i.e. $f(x) = f(-x)$ for all $x \in \mathbb{R}$), and assume that $\frac{f'(x)}{x}$ is decreasing over the interval $[0,\infty)$. Then the function*

$$f(\tilde{x}) + \frac{f'(\tilde{x})}{2\tilde{x}}(x^2 - \tilde{x}^2) \tag{41}$$

majorizes f at $\tilde{x}$.

Proof: See [46, Theorem 4.5]. ■

Lemma 3. *Let $f(x) = -\log(1+\mu-\frac{\mu}{1+\eta x^2})$ for some $\mu,\eta > 0$. Then for all $x, \tilde{x} \in \mathbb{R}$ we have that*

$$f(x) \leq f(\tilde{x}) + \frac{\eta}{1+\eta\tilde{x}^2}(x^2-\tilde{x}^2) - \frac{2\eta\tilde{x}(1+\mu)}{1+\eta(1+\mu)\tilde{x}^2}(x-\tilde{x}) + \eta(1+\mu)(x-\tilde{x})^2.$$

Proof: We can rewrite $f(x)$ as

$$f(x) = \log(1+\eta x^2) - \log(1+\eta x^2 + \mu\eta x^2). \tag{42}$$

The first term satisfies the conditions in Lemma 2 and hence its majorizer is given by

$$\log(1+\eta x^2) \leq \log(1+\eta\tilde{x}^2) + \frac{\eta}{1+\eta\tilde{x}^2}(x^2-\tilde{x}^2).$$

Moreover, let $f_2(x) = -\log(1+\eta x^2 + \mu\eta x^2)$. It is straightforward to verify that

$$f_2''(x) = \frac{4\eta^2(1+\mu)^2x^2 - 2\eta(1+\mu)(1+\eta x^2(1+\mu))}{1+\eta^2(1+\mu)^2x^4 + 2\eta(1+\mu)x^2} \leq 2\eta(1+\mu)\frac{2\eta(1+\mu)x^2}{1+2\eta(1+\mu)x^2} \leq 2\eta(1+\mu).$$

Consequently, $f_2(x)$ can be majorized using Lemma 1, and hence the proof is concluded. ■

- **Bhattacharyya distance:** For $\mathcal{I} = \mathcal{B}$, substituting $\{\lambda_k\}$ of (38) into the objective function of (37) leads to the following expression for the objective function:

$$\sum_{k=1}^{N_r}\left[\log\left(1+0.5\gamma_k - 0.5\frac{\gamma_k}{1+\beta_k \mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}}\right) + \frac{0.5}{1+\lambda_k^{(l)}}\left(\frac{\gamma_k}{1+\beta_k \mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}}\right)\right] \tag{43}$$

where

$$\lambda_k^{(l)} \;=\; \gamma_k - \frac{\gamma_k}{1+\beta_k y_k^{(l)}}. \tag{44}$$

A minorizer of the logarithmic term can be obtained immediately by employing Lemma 3 with $x_k = \sqrt{\mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}}$, $\mu = 0.5\gamma_k$, and $\eta = \beta_k$. To deal with the expression $\frac{\gamma_k}{1+\beta_k \mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}}$ in (43) conveniently, we use the convexity of the function $\frac{1}{1+\beta x}$ for $\beta > 0$ which implies

$$\frac{1}{1+\beta x} \geq \frac{1}{1+\beta\tilde{x}} - \frac{\beta}{(1+\beta\tilde{x})^2}(x-\tilde{x}), \qquad \forall\; x, \tilde{x}. \tag{45}$$

As a result, a minorizer of $\frac{\gamma_k}{1+\beta_k \mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}}$ can be obtained by considering the above inequality for $x_k = \mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}$ and $\beta = \beta_k$. Furthermore, by replacing the summation terms in (43) for each k with the obtained minorizers (using Lemma 3 and eq. (45)) and removing the constants, the criterion in (43) turns to:

$$\begin{aligned}\sum_{k=1}^{N_r}\Bigg[-\left(\frac{\beta_k}{1+\beta_k y_k^{(l)}} + \beta_k(1+0.5\gamma_k) + \frac{0.5\gamma_k}{1+\lambda_k^{(l)}}\,\frac{\beta_k}{(1+\beta_k y_k^{(l)})^2}\right)\mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a} \\ + \left(\frac{2\beta_k(1+0.5\gamma_k)\sqrt{y_k^{(l)}}}{1+\beta_k y_k^{(l)}(1+0.5\gamma_k)} + 2\beta_k(1+0.5\gamma_k)\sqrt{y_k^{(l)}}\right)\sqrt{\mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}}\Bigg]\end{aligned} \tag{46}$$

where

$$y_k^{(l)} = (\mathbf{a}^{(l)})^H\mathbf{M}_k^{-1}\mathbf{a}^{(l)}. \tag{47}$$

Yet, due to the non-concavity of the terms $\left\{\sqrt{\mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}}\right\}$, dealing with the maximization of the criterion in (46) appears to be complicated. However, $\sqrt{\mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}}$ can be minorized using its supporting hyperplane at any given $\tilde{\mathbf{a}}$; more precisely,

$$\sqrt{\mathbf{a}^H\mathbf{M}_k^{-1}\mathbf{a}} \geq \sqrt{\tilde{\mathbf{a}}^H\mathbf{M}_k^{-1}\tilde{\mathbf{a}}} + \Re\left(\frac{\tilde{\mathbf{a}}^H\mathbf{M}_k^{-1}}{\sqrt{\tilde{\mathbf{a}}^H\mathbf{M}_k^{-1}\tilde{\mathbf{a}}}}(\mathbf{a}-\tilde{\mathbf{a}})\right). \tag{48}$$

The above inequality holds true due to the convexity of the function $h(\mathbf{x}) = \|\mathbf{x}\|_2$ and the fact that the gradient of $h(\mathbf{x})$ is given by $\nabla h(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}$. Ultimately, by using eq. (46) and (48) as well as removing the constants, the optimization problem associated with the $(l+1)^{th}$ iteration of Sv-MaMi for $\mathcal{I} = \mathcal{B}$ is as follows:

$$\begin{aligned} \min_{\mathbf{a}} \quad & \mathbf{a}^H \left(\sum_{k=1}^{N_r} \phi_{k,\mathcal{B}}^{(l)} \mathbf{M}_k^{-1} \right) \mathbf{a} - \Re \left(\left(\sum_{k=1}^{N_r} \mathbf{d}_{k,\mathcal{B}}^{(l)} \right)^H \mathbf{a} \right) \\ \text{subject to} \quad & \|\mathbf{a}\|_2^2 \leqslant e, \end{aligned} \tag{49}$$

where

$$\begin{aligned} \phi_{k,\mathcal{B}}^{(l)} &\triangleq \left(\frac{\beta_k}{1+\beta_k y_k^{(l)}} + \beta_k(1+0.5\gamma_k) + \frac{0.5\gamma_k}{1+\lambda_k^{(l)}} \, \frac{\beta_k}{(1+\beta_k y_k^{(l)})^2} \right) \\ \mathbf{d}_{k,\mathcal{B}}^{(l)} &\triangleq \left(\frac{2\beta_k(1+0.5\gamma_k)}{1+\beta_k y_k^{(l)}(1+0.5\gamma_k)} + 2\beta_k(1+0.5\gamma_k) \right) \mathbf{M}_k^{-1}\mathbf{a}^{(l)}. \end{aligned} \tag{50}$$

Note that as $\phi_{k,\mathcal{B}}^{(l)} > 0$ and $\mathbf{M}_k \succ \mathbf{0}$, the above problem is a convex QCQP.

- **KL-Divergence:** In the case of $\mathcal{I} = \mathcal{D}$, using (37) and (38) (and dropping the constants) leads to the following expression for the corresponding objective function:

$$\sum_{k=1}^{N_r} \left[\log\left(1+\gamma_k - \frac{\gamma_k}{1+\beta_k \mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}}\right) + \left(\frac{1}{1+\lambda_k^{(l)}}\right)^2 \left(\frac{\gamma_k}{1+\beta_k \mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}}\right) \right]. \tag{51}$$

The logarithmic term in (51) can be handled via Lemma 3 by setting $x_k = \sqrt{\mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}}, \mu = \gamma_k$, and $\eta = \beta_k$. Moreover, the expression $\frac{\gamma_k}{1+\beta_k \mathbf{a}^H \mathbf{M}_k^{-1}\mathbf{a}}$ can be minorized using (48). Consequently, using a similar approach as in the case of the Bhattacharrya distance, the optimization problem associated with the $(l+1)^{th}$ iteration of Sv-MaMi for $\mathcal{I} = \mathcal{D}$ is given by:

$$\begin{aligned} \min_{\mathbf{a}} \quad & \mathbf{a}^H \left(\sum_{k=1}^{N_r} \phi_{k,\mathcal{D}}^{(l)} \mathbf{M}_k^{-1} \right) \mathbf{a} - \Re \left(\left(\sum_{k=1}^{N_r} \mathbf{d}_{k,\mathcal{D}}^{(l)} \right)^H \mathbf{a} \right) \\ \text{subject to} \quad & \|\mathbf{a}\|_2^2 \leqslant e, \end{aligned} \tag{52}$$

where

$$\begin{aligned} \phi_{k,\mathcal{D}}^{(l)} &\triangleq \left(\frac{\beta_k}{1+\beta_k y_k^{(l)}} + \beta_k(1+\gamma_k) + \left(\frac{\gamma_k}{1+\lambda_k^{(l)}}\right)^2 \left(\frac{\beta_k}{(1+\beta_k y_k^{(l)})^2} \right) \right) \\ \mathbf{d}_{k,\mathcal{D}}^{(l)} &\triangleq \left(\frac{2\beta_k(1+\gamma_k)}{1+\beta_k y_k^{(l)}(1+\gamma_k)} + 2\beta_k(1+\gamma_k) \right) \mathbf{M}_k^{-1}\mathbf{a}^{(l)}. \end{aligned}$$

- **J-Divergence:** In this case, (37) boils down to the following non-convex optimization problem:

$$\begin{aligned} \min_{\mathbf{a}} \quad & \sum_{k=1}^{N_r} \gamma_k w_k^{(l)} \left(\frac{1}{1+\beta_k \mathbf{a}^H \mathbf{M}_k^{-1} \mathbf{a}} \right) \\ \text{subject to} \quad & \|\mathbf{a}\|_2^2 \leq e, \end{aligned} \tag{53}$$

with $w_k^{(l)} = 1 - \frac{1}{(1+\lambda_k^{(l)})^2} > 0$. Note that in contrast to the case of Bhattacharyya distance and KL-divergence (see eqs. (43) and (51)), the expression $\frac{1}{1+\beta_k \mathbf{a}^H \mathbf{M}_k^{-1} \mathbf{a}}$ appears in a *minimization* problem. We consider a majorization of the function $\frac{1}{1+\eta x^2}$ (note that $\frac{1}{1+\eta x^2} = \frac{1}{1+\beta_k \mathbf{a}^H \mathbf{M}_k^{-1} \mathbf{a}}$ when $x_k = \sqrt{\mathbf{a}^H \mathbf{M}_k^{-1} \mathbf{a}}$, $\eta = \beta_k$). One can derive a majorizer for the aforementioned function via Lemma 1, viz.

$$\frac{1}{1+\eta x^2} \leq \frac{1}{1+\eta \tilde{x}^2} - \frac{2\eta\tilde{x}}{(1+\eta\tilde{x}^2)^2}(x-\tilde{x}) + \frac{1}{2}U(x-\tilde{x})^2. \tag{54}$$

Note that we have

$$\frac{d^2}{dx^2}\left(\frac{1}{1+\eta x^2}\right) = \frac{6\eta^2 x^2 - 2\eta}{(1+\eta x^2)^3} \leq 6\eta$$

which implies that (54) holds true for $U = 6\eta$. Therefore, by minorizing $\sqrt{\mathbf{a}^H \mathbf{M}_k^{-1} \mathbf{a}}$ using (48), the following QCQP is obtained for the $(l+1)^{th}$ iteration of Sv-MaMi for $\mathcal{I} = \mathcal{J}$:

$$\begin{aligned} \min_{\mathbf{a}} \quad & \mathbf{a}^H \left(\sum_{k=1}^{N_r} \phi_{k,\mathcal{J}}^{(l)} \mathbf{M}_k^{-1} \right) \mathbf{a} - \Re\left(\left(\sum_{k=1}^{N_r} \mathbf{d}_{k,\mathcal{J}}^{(l)} \right)^H \mathbf{a} \right) \\ \text{subject to} \quad & \|\mathbf{a}\|_2^2 \leqslant e, \end{aligned} \tag{55}$$

where

$$\begin{aligned} \phi_{k,\mathcal{J}}^{(l)} &\triangleq 3\beta_k \gamma_k w_k^{(l)} \\ \mathbf{d}_{k,\mathcal{J}}^{(l)} &\triangleq \left(\frac{2\beta_k \gamma_k w_k^{(l)}}{(1+\beta_k y_k^{(l)})^2} + 6\beta_k \right) \mathbf{M}_k^{-1} \mathbf{a}^{(l)}. \end{aligned} \tag{56}$$

- **Mutual Information:** The derivation of the QCQP corresponding to $\mathcal{I} = \mathcal{M}$ is straightforward. In particular, using Lemma 3 as well as (48) we obtain the following QCQP:

$$\begin{aligned} \min_{\mathbf{a}} \quad & \mathbf{a}^H \left(\sum_{k=1}^{N_r} \phi_{k,\mathcal{M}}^{(l)} \mathbf{M}_k^{-1} \right) \mathbf{a} - \Re\left(\left(\sum_{k=1}^{N_r} \mathbf{d}_{k,\mathcal{M}}^{(l)} \right)^H \mathbf{a} \right) \\ \text{subject to} \quad & \|\mathbf{a}\|_2^2 \leqslant e, \end{aligned} \tag{57}$$

where

$$\begin{aligned} \phi_{k,\mathcal{M}}^{(l)} &\triangleq \left(\frac{\beta_k}{1+\beta_k y_k^{(l)}} + \beta_k(1+\gamma_k) \right) \\ \mathbf{d}_{k,\mathcal{M}}^{(l)} &\triangleq \left(\frac{2\beta_k(1+\gamma_k)}{1+\beta_k y_k^{(l)}(1+\gamma_k)} + 2\beta_k(1+\gamma_k) \right) \mathbf{M}_k^{-1} \mathbf{a}^{(l)}. \end{aligned} \tag{58}$$

Table I summarizes the steps of Sv-MaMi. Note that the convex QCQP of the first step can be solved very efficiently (see e.g. [47]). Moreover, the derivations of Sv-MaMi algorithm can be extended to tackle code design problems in which a PAR-constrained code is required. Such an extension is discussed in Section VI.

TABLE I

THE SV-MAMI ALGORITHM FOR $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$

Step 0: Initialize $\mathbf{a}$ with a random vector in $\mathbb{C}^N$ and set the iteration number l to 0.
Step 1: Solve the QCQP problem in (34) to obtain $\mathbf{a}^{(l+1)}$; set $l \leftarrow l+1$.
Step 2: Compute $\phi_{k,\mathcal{I}}^{(l)}$ and $\mathbf{d}_{k,\mathcal{I}}^{(l)}$ corresponding to the metric $\mathcal{I}$.
Step 3: Repeat steps 1 and 2 until a pre-defined stop criterion is satisfied, e.g. $\|\mathbf{a}^{(l+1)} - \mathbf{a}^{(l)}\|_2 \leq \xi$ for some $\xi > 0$.

Remark 3 (Saturation Phenomenon): It might be of interest to investigate the behavior of the considered information-theoretic criteria when the transmit energy e grows large. Let $\bar{\mathbf{a}}$ represent the unit-norm version of $\mathbf{a}$ (i.e. $\mathbf{a} = \sqrt{e}\ \bar{\mathbf{a}}$) and note that:

$$\lim_{e \to \infty} q_{\mathcal{I}} \left(\gamma_k - \frac{\gamma_k}{1 + e\beta_k \bar{\mathbf{a}}^H \mathbf{M}_k^{-1} \bar{\mathbf{a}}} \right) = q_{\mathcal{I}}(\gamma_k). \tag{59}$$

In light of the above equality, one can observe that all information-theoretic criteria $q_{\mathcal{I}}$ for $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$ tend to constant values in $\mathbb{R}_+$ as $e \to \infty$. We refer to this behavior of the considered metrics as the saturation phenomenon, meaning that for sufficiently large values of the transmit energy e, the performance improvement obtained by choosing $\mathbf{a}$ or by increasing e is negligible. Interestingly, it might still be reasonable to increase the transmit energy of the system. Indeed our previous arguments rely on the fact that a fixed radar cell is considered; however, increasing e extends the detection range (or coverage) of the system. ■

V. CODE DESIGN USING MAMI AND RELAXATION

In this section, we propose another algorithm based on MaMi to tackle the optimization problems formulated in (25). The suggested algorithm (which we call Re-MaMi) employs a relaxation of the rank constraint on the code matrix $\mathbf{A} = \mathbf{a}\mathbf{a}^H$ such that each iteration of MaMi can be handled as a convex optimization problem. In particular, we have the following result:

Theorem 2. *(Re-MaMi algorithm) The solution code matrix $\mathbf{A} = \mathbf{A}_\star$ (with relaxed rank constraint) can be obtained iteratively by solving the following convex problem (at the $(l+1)^{th}$ iteration):*

$$\begin{aligned} \max_{\mathbf{A}} \quad & \sum_{k=1}^{N_r} \left[f_{\mathcal{I}} \left(N\gamma_k - \gamma_k \operatorname{tr}\{(\mathbf{M}_k + \beta_k \mathbf{A})^{-1}\mathbf{M}_k\} \right) + h_{k,\mathcal{I}}^{(l)}(\mathbf{A}) \right] \\ \textit{subject to} \quad & \operatorname{tr}\{\mathbf{A}\} \leqslant e \\ & \mathbf{A} \succeq \mathbf{0}, \end{aligned} \tag{60}$$

where $h_{k,\mathcal{I}}^{(l)}(\mathbf{A})$ denotes a concave function of $\mathbf{A}$, for $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$, that will be given in explicit form below.

In what follows, we present a proof of Theorem 2 and then discuss the synthesis of optimized $\mathbf{a}_\star$ from the obtained $\mathbf{A}_\star$. First note that using matrix inversion lemma, λ_k can be rewritten as

$$\begin{aligned} \lambda_k &= \sigma_k^2 \mathbf{a}^H (\sigma_{c,k}^2 \mathbf{a}\mathbf{a}^H + \mathbf{M}_k)^{-1}\mathbf{a} = \sigma_k^2 \operatorname{tr}\{(\sigma_{c,k}^2 \mathbf{a}\mathbf{a}^H + \mathbf{M}_k)^{-1}\mathbf{a}\mathbf{a}^H\} \\ &= \sigma_k^2 \operatorname{tr}\{(\sigma_{c,k}^2 \mathbf{A} + \mathbf{M}_k)^{-1}\mathbf{A}\} = \frac{\sigma_k^2}{\sigma_{c,k}^2} \operatorname{tr}\{\mathbf{I} - (\mathbf{I} + \sigma_{c,k}^2 \mathbf{M}_k^{-1}\mathbf{A})^{-1}\} \\ &= N\frac{\sigma_k^2}{\sigma_{c,k}^2} - \frac{\sigma_k^2}{\sigma_{c,k}^2} \operatorname{tr}\{(\mathbf{M}_k + \sigma_{c,k}^2 \mathbf{A})^{-1}\mathbf{M}_k\} \triangleq N\gamma_k - \gamma_k \operatorname{tr}\{(\mathbf{M}_k + \beta_k \mathbf{A})^{-1}\mathbf{M}_k\} \end{aligned} \tag{61}$$

where $\mathbf{A} = \mathbf{a}\mathbf{a}^H$. As a result, using (61) and (36), the optimal code matrix $\mathbf{A} = \mathbf{A}_\star$ can be obtained iteratively via solving the following optimization problem at the $(l+1)^{th}$ iteration:

$$\max_{\mathbf{A},\lambda_k} \quad \sum_{k=1}^{N_r} f_{\mathcal{I}}(\lambda_k) + g'_{\mathcal{I}}\left(\lambda_k^{(l)}\right) \lambda_k \tag{62}$$

$$\text{subject to} \quad \lambda_k = N\gamma_k - \gamma_k \operatorname{tr}\left\{(\mathbf{M}_k + \beta_k \mathbf{A})^{-1}\mathbf{M}_k\right\} \tag{63}$$

$$\operatorname{tr}\{\mathbf{A}\} \leqslant e \tag{64}$$

$$\mathbf{A} \succeq \mathbf{0} \tag{65}$$

$$\operatorname{rank}(\mathbf{A}) = 1. \tag{66}$$

Note that the above problem is non-convex due to the non-affine equality constraints in (63) and (66). Hereafter, we relax the rank-one constraint (66). Moreover, when the term $\left\{ g'_{\mathcal{I}}\left(\lambda_k^{(l)}\right) \lambda_k \right\}$ is not concave (with respect to $\mathbf{A}$), a further minorization will be needed in order to make the problem convex. Let $h_{k,\mathcal{I}}^{(l)}(\mathbf{A})$ (to be discussed shortly) denote a concave function that minorizes $g'_{\mathcal{I}}\left(\lambda_k^{(l)}\right) \lambda_k$ (we let $h_{k,\mathcal{I}}^{(l)} = g'_{\mathcal{I}}\left(\lambda_k^{(l)}\right) \lambda_k$ when $g'_{\mathcal{I}}\left(\lambda_k^{(l)}\right) \lambda_k$ is concave itself).

Remark 4: Note that for $\mathbf{A} \succeq \mathbf{0}$ of rank δ, there exists a $\mathbf{V} \in \mathbb{C}^{N\times\delta}$ such that $\mathbf{A} = \mathbf{V}\mathbf{V}^H$. As a result, considering (61) we have

$$\begin{aligned} N\gamma_k - \gamma_k \operatorname{tr}\{(\mathbf{M}_k + \beta_k\mathbf{A})^{-1}\mathbf{M}_k\} &= \sigma_k^2 \operatorname{tr}\{(\sigma_{c,k}^2\mathbf{A} + \mathbf{M}_k)^{-1}\mathbf{A}\} \\ &= \sigma_k^2 \operatorname{tr}\{\mathbf{V}^H(\sigma_{c,k}^2\mathbf{A} + \mathbf{M}_k)^{-1}\mathbf{V}\} > 0 \end{aligned}$$

which implies that the argument of the function $f_{\mathcal{I}}(.)$ in (62) remains positive even in the case in which no rank constraint on $\mathbf{A}$ is imposed. Moreover, note that $\operatorname{tr}\{(\mathbf{M}_k + \beta_k\mathbf{A})^{-1}\mathbf{M}_k\}$ is a convex function of $\mathbf{A}$. Consequently, $f_{\mathcal{I}}\left(N\gamma_k - \gamma_k \operatorname{tr}\{(\mathbf{M}_k + \beta_k\mathbf{A})^{-1}\mathbf{M}_k\}\right)$ is a concave function of $\mathbf{A}$ as $f_{\mathcal{I}}(.)$ is an increasing function (for all $\mathcal{I}$). ■

Selecting a suitable function $h_{k,\mathcal{I}}^{(l)}(\mathbf{A})$ depends on the code design metric:

- **Bhattacharyya distance:** By substituting λ_k of (63) as well as $g'_{\mathcal{B}}\left(\lambda_k^{(l)}\right)$, the objective function in (62) for $\mathcal{I} = \mathcal{B}$ can be rewritten (by omitting constants) as

$$\begin{aligned} \sum_{k=1}^{N_r} \Big[&\log\left(1 + 0.5N\gamma_k - 0.5\gamma_k \operatorname{tr}\{(\mathbf{M}_k + \beta_k\mathbf{A})^{-1}\mathbf{M}_k\}\right) \\ &+ \frac{0.5\gamma_k}{1+\lambda_k^{(l)}} \operatorname{tr}\{(\mathbf{M}_k + \beta_k\mathbf{A})^{-1}\mathbf{M}_k\}\Big] \end{aligned}$$

where $\lambda_k^{(l)} = N\gamma_k - \gamma_k \operatorname{tr}\{(\mathbf{M}_k + \beta_k\mathbf{A}^{(l)})^{-1}\mathbf{M}_k\}$ with $\mathbf{A}^{(l)}$ being the $\mathbf{A}$ obtained at the l^{th} iteration. As mentioned in Remark 4 the logarithmic term is concave; however, the second term is convex with respect to (w.r.t.) $\mathbf{A}$, and hence a minorization is needed to tackle the problem.

Lemma 4. *Let $h(\mathbf{A}) = \operatorname{tr}\{(\mathbf{M}_k + \beta_k\mathbf{A})^{-1}\mathbf{M}_k\}$. A minorizer of $h(\mathbf{A})$ at $\mathbf{A} = \tilde{\mathbf{A}}$ is given by*

$$\tilde{h}(\mathbf{A}) = \operatorname{tr}\{(\mathbf{M}_k + \beta_k\tilde{\mathbf{A}})^{-1}\mathbf{M}_k\} - \beta_k \operatorname{tr}\{(\mathbf{M}_k + \beta_k\tilde{\mathbf{A}})^{-1}\mathbf{M}_k(\mathbf{M}_k + \beta_k\tilde{\mathbf{A}})^{-1}(\mathbf{A} - \tilde{\mathbf{A}})\}.$$

Proof: See Appendix C. ■

By applying Lemma 4 to (67), we can recast the maximization step at the $(l+1)^{th}$ iteration of Re-MaMi for $\mathcal{I} = \mathcal{B}$ as follows:

$$\begin{aligned} \max_{\mathbf{A}} \quad & \sum_{k=1}^{N_r} \left[\log\left(1 + 0.5N\gamma_k - 0.5\gamma_k \operatorname{tr}\{(\mathbf{M}_k + \beta_k\mathbf{A})^{-1}\mathbf{M}_k\}\right) + h_{k,\mathcal{B}}^{(l)}(\mathbf{A})\right] \\ \text{subject to} \quad & \operatorname{tr}\{\mathbf{A}\} \leqslant e \\ & \mathbf{A} \succeq \mathbf{0}, \end{aligned} \tag{67}$$

where

$$h_{k,\mathcal{B}}^{(l)}(\mathbf{A}) \triangleq -\operatorname{tr}\left\{\mathbf{F}_{k,\mathcal{B}}^{(l)}\mathbf{A}\right\}, \tag{68}$$
$$\mathbf{F}_{k,\mathcal{B}}^{(l)} \triangleq \frac{0.5\ \gamma_k\beta_k}{1+\lambda_k^{(l)}}\left(\mathbf{M}_k+\beta_k\mathbf{A}^{(l)}\right)^{-1}\mathbf{M}_k\left(\mathbf{M}_k+\beta_k\mathbf{A}^{(l)}\right)^{-1}.$$

- **KL-divergence:** By substituting λ_k and $g'_{\mathcal{D}}(\lambda_k^{(l)})$ in (62), it can be easily verified that (62) includes the expression $\operatorname{tr}\{(\mathbf{M}_k+\beta_k\mathbf{A})^{-1}\mathbf{M}_k\}$ with positive sign. Therefore, similar to the case of Bhattacharyya distance, the following convex problem can be derived (using Lemma 4) at the $(l+1)^{th}$ iteration of Re-MaMi for $\mathcal{I}=\mathcal{D}$:

$$\begin{aligned} \max_{\mathbf{A}} \quad & \sum_{k=1}^{N_r}\Big[\log\big(1+N\gamma_k-\gamma_k\operatorname{tr}\{(\mathbf{M}_k+\beta_k\mathbf{A})^{-1}\mathbf{M}_k\}\big)+h_{k,\mathcal{D}}^{(l)}(\mathbf{A})\Big] \\ \text{subject to} \quad & \operatorname{tr}\{\mathbf{A}\}\leqslant e \\ & \mathbf{A}\succeq\mathbf{0}, \end{aligned} \tag{69}$$

where

$$h_{k,\mathcal{D}}^{(l)}(\mathbf{A}) \triangleq -\operatorname{tr}\left\{\mathbf{F}_{k,\mathcal{D}}^{(l)}\mathbf{A}\right\},$$
$$\mathbf{F}_{k,\mathcal{D}}^{(l)} \triangleq \frac{\gamma_k\beta_k}{\left(1+\lambda_k^{(l)}\right)^2}\left(\mathbf{M}_k+\beta_k\mathbf{A}^{(l)}\right)^{-1}\mathbf{M}_k\left(\mathbf{M}_k+\beta_k\mathbf{A}^{(l)}\right)^{-1}. \tag{70}$$

- **J-divergence:** For the case of $\mathcal{I}=\mathcal{J}$, the relaxed version of the maximization in (62)-(66) is equivalent to the optimization problem:

$$\begin{aligned} \min_{\mathbf{A}} \quad & \sum_{k=1}^{N_r}\gamma_k w_k^{(l)}\operatorname{tr}\left\{(\mathbf{M}_k+\beta_k\mathbf{A})^{-1}\mathbf{M}_k\right\} \\ \text{subject to} \quad & \operatorname{tr}\{\mathbf{A}\}\leqslant e \\ & \mathbf{A}\succeq\mathbf{0}, \end{aligned} \tag{71}$$

where $w_k^{(l)}=1-(\frac{1}{1+\lambda_k^{(l)}})^2$ and $h_{k,\mathcal{J}}^{(l)}=-\gamma_k w_k^{(l)}\operatorname{tr}\left\{(\mathbf{M}_k+\beta_k\mathbf{A})^{-1}\mathbf{M}_k\right\}$.

Note that Remark 4 ensures that $w_k^{(l)}>0$ for all (k,l), and hence (71) is a convex optimization problem due to the convexity of $\operatorname{tr}\{\mathbf{X}^{-1}\}$ w.r.t. $\mathbf{X}\succ\mathbf{0}$. Note also that the optimization problem in (71) can be recast as a semi-definite program (SDP) by considering an SDP representation of the $\operatorname{tr}\{\mathbf{X}^{-1}\}$ minimization (see e.g. [48]).

- **Mutual information:** Using the relaxation of the rank-one constraint for the case of $\mathcal{I}=\mathcal{M}$, one obtains the following form of the optimization problem in (62)-(66):

$$\begin{aligned} \max_{\mathbf{A}} \quad & \sum_{k=1}^{N_r} \log\left(1 + N\gamma_k - \gamma_k \operatorname{tr}\{(\mathbf{M}_k + \beta_k \mathbf{A})^{-1}\mathbf{M}_k\}\right) \\ \text{subject to} \quad & \operatorname{tr}\{\mathbf{A}\} \leqslant e \\ & \mathbf{A} \succeq \mathbf{0}. \end{aligned} \tag{72}$$

Note that the above relaxed version of the optimization in (62)-(66) is a convex problem that can be solved in one iteration of Re-MaMi (as no majorization is required, i.e. $h_{k,\mathcal{M}}^{(l)}(\mathbf{A}) = 0$).

We end this section by discussing the synthesis stage required for Re-MaMi. Once the proposed Re-MaMi algorithm converges to $\mathbf{A}_\star$, the corresponding code $\mathbf{a}_\star$ can be obtained as follows. If $\operatorname{rank}(\mathbf{A}_\star) = 1$, the local optimum obtained for the relaxed problem in (60) yields a local optimum of (25) via $\mathbf{A}_\star = \mathbf{a}_\star \mathbf{a}_\star^H$. Otherwise, a synthesis loss is unavoidable due to the rank of $\mathbf{A}_\star$ being larger than 1. The rank behavior of the matrix obtained from the relaxed problem and the associated rank-one approximations have been discussed in the literature particularly for semi-definite relaxations (see e.g. [49], [50] and references therein). Least-squares (LS) synthesis is a common approach to synthesize the optimized codes [49]. The LS criterion can be formulated as:

$$\min_{\mathbf{a}} \ \|\mathbf{A}_\star - \mathbf{a}\mathbf{a}^H\|_F^2 \quad \text{subject to} \quad \|\mathbf{a}\|_2^2 = e. \tag{73}$$

The solution to the above problem is simply given by $\sqrt{e}\widetilde{\mathbf{b}}$ where $\widetilde{\mathbf{b}}$ is the principal eigenvector of $\mathbf{A}_\star$. Inspired by the randomization technique in the literature (see e.g. [49] and the references therein), here we employ randomization as an alternative approach of code synthesis. In the randomization technique, several feasible random vectors $\{\mathbf{a}_m\}_{m=1}^L$ are generated (e.g. according to the distribution $\mathcal{CN}(\mathbf{0}, \mathbf{A}_\star)$) and $\mathbf{a}_\star$ is obtained as

$$\mathbf{a}_\star = \arg\max_m \{\widetilde{q}_{\mathcal{I}}(\mathbf{a}_m)\} \tag{74}$$

where $\widetilde{q}_{\mathcal{I}}(\cdot)$ denotes the objective function in (60).

The steps of Re-MaMi algorithm are summarized in Table II. Note that the first step of Re-MaMi (for all $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$) contains a convex problem which can be solved efficiently via interior point methods [51]. Modification of Re-MaMi to obtain optimized codes under a PAR constraint is discussed in the next section.

VI. Extensions of the Design Methods

In this section we provide two extensions of the previous design methods to PAR-constrained codes and the case of multiple transmitters. In order to use the power resources efficiently and to avoid non-linear

TABLE II

THE RE-MAMI ALGORITHM FOR $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$

Step 0: Initialize $\mathbf{A}$ with a random matrix in $\mathbb{C}^{N\times N}$ and set the iteration number l to 0.
Step 1: Solve the convex problem in (60) to obtain $\mathbf{A}^{(l+1)}$; set $l \leftarrow l+1$.
Step 2: Compute $h_{k,\mathcal{I}}^{(l)}(\mathbf{A})$ corresponding to the metric $\mathcal{I}$.
Step 3: Repeat steps 1 and 2 until a pre-defined stop criterion is satisfied, e.g. $\|\mathbf{A}^{(l+1)} - \mathbf{A}^{(l)}\|_F \leq \xi$ for some $\xi > 0$.
Step 4 (*Synthesis stage*): Synthesize the optimized code $\mathbf{a}_\star$ using the approaches in (73) or (74).

effects at the transmitter, codes with low PAR values are of practical interest in many applications [52] [53]. To the best of our knowledge, no study of code design with PAR constraints using information-theoretic criteria was conducted prior to this work. This section also includes the extension of the design methods to deal with the case of multiple transmitters with orthogonal transmission.

A. PAR-constrained code design

In this subsection, we extend the derivations of Sv-MaMi and Re-MaMi for code design with an arbitrary PAR constraint, viz.

$$\mathrm{PAR}(\mathbf{a}) = \frac{\max\limits_n \{|a_n|^2\}}{\frac{1}{N}\|\mathbf{a}\|_2^2} \leq \zeta. \tag{75}$$

For Sv-MaMi the PAR constrained problem that must be solved is:

$$\begin{aligned} \min_{\mathbf{a}} \quad & \mathbf{a}^H \left(\sum_{k=1}^{N_r} \phi_{k,\mathcal{I}}^{(l)} \mathbf{M}_k^{-1} \right) \mathbf{a} - \Re \left(\left(\sum_{k=1}^{N_r} \mathbf{d}_{k,\mathcal{I}}^{(l)} \right)^H \mathbf{a} \right) \\ \text{subject to} \quad & \max_{n=1,\ldots,N} \{|a_n|^2\} \leq \zeta \\ & \|\mathbf{a}\|_2^2 = N. \end{aligned} \tag{76}$$

For Re-MaMi, one can consider the PAR constraint in the synthesis stage, which for LS synthesis leads to the following optimization problem:

$$\begin{aligned} \max_{\mathbf{a}} \quad & \mathbf{a}^H \mathbf{A} \mathbf{a} \\ \text{subject to} \quad & \max_{n=1,\ldots,N} \{|a_n|^2\} \leq \zeta \\ & \|\mathbf{a}\|_2^2 = N. \end{aligned} \tag{77}$$

Note that the QCQPs in (76) and (77) are non-convex optimization problems and known to be NP-hard [52]. Also note that the problem in (76) can be recast in a form similar to (77), viz.

$$\begin{aligned} \min_{\mathbf{a}} \quad & \widehat{\mathbf{a}}^H \mathbf{J}\, \widehat{\mathbf{a}} \\ \text{subject to} \quad & \max_{n=1,\cdots,N} \{|a_n|^2\} \leq \zeta \\ & \|\mathbf{a}\|_2^2 = N \end{aligned} \tag{78}$$

where $\widehat{\mathbf{a}} = [\mathbf{a} \quad 1]^T$, and

$$\mathbf{J} = \begin{bmatrix} \left(\sum_{k=1}^{N_r} \phi_{k,\mathcal{I}}^{(l)} \mathbf{M}_k^{-1}\right) & -0.5\left(\sum_{k=1}^{N_r} \mathbf{d}_{k,\mathcal{I}}^{(l)}\right) \\ -0.5\left(\sum_{k=1}^{N_r} \mathbf{d}_{k,\mathcal{I}}^{(l)}\right)^H & 0 \end{bmatrix}.$$

In what follows, we will explain how to solve (76) but, of course, (77) can be tackled in the same way. Let $\mathbf{K} = \mu \mathbf{I}_{N+1} - \mathbf{J}$ for any $\mu > \mu_{max}(\mathbf{J})$. Next observe that the problem in (78) is equivalent to:

$$\begin{aligned} \max_{\mathbf{a}} \quad & \widehat{\mathbf{a}}^H \mathbf{K}\, \widehat{\mathbf{a}} \\ \text{subject to} \quad & \max_{n=1,\cdots,N} \{|a_n|^2\} \leq \zeta \\ & \|\mathbf{a}\|_2^2 = N. \end{aligned} \tag{79}$$

The above problem can be tackled using the power-method in [54]. More precisely, the code vector $\mathbf{a}$ at the $(p+1)^{th}$ iteration can be obtained from $\mathbf{a}^{(p)}$, via solving the optimization problem

$$\begin{aligned} \max_{\mathbf{a}^{(p+1)}} \quad & \|\mathbf{a}^{(p+1)} - \breve{\mathbf{a}}^{(p)}\| \\ \text{subject to} \quad & \max_{n=1,\cdots,N} \{|a_n^{(p+1)}|^2\} \leq \zeta \\ & \|\mathbf{a}^{(p+1)}\|_2^2 = N \end{aligned} \tag{80}$$

where $\breve{\mathbf{a}}^{(p)}$ represents the vector containing the first N entries of $\mathbf{K}\, \widehat{\mathbf{a}}^{(p)}$. The optimization problem (80) is a "nearest-vector" problem with PAR constraint. Such PAR constrained problems can be tackled using an algorithm proposed in [55]. Note that the codes obtained as above can be scaled to fit any desired level of transmit energy as a scaling does not affect the PAR metric (see (75)). We refer the interested reader to [52] for using the randomization technique when a PAR constraint is imposed.

B. The case of multiple transmitters

Here we discuss the extension of the design problem to the case of multiple transmit antennas that emit orthogonal signals. Let $\widetilde{s}_m(t)$ and $\mathbf{a}_m$ denote the passband version and associated code vector of the m^{th} transmit signal, respectively. Assume that $\{\widetilde{s}_m(t)\}_{m=1}^{N_t}$ are well-separated in the frequency domain such that the signal echoes corresponding to each transmitter can be extracted at the k^{th} receiver. Then, the discrete-time signal at the k^{th} receiver due to the m^{th} transmitter can be expressed as

$$\mathbf{r}_{k,m} = \alpha_{k,m}\mathbf{a}_m + \widetilde{\rho}_{k,m}\mathbf{a}_m + \mathbf{w}_{k,m},\ m = 1, 2, ..., N_t; k = 1, 2, ..., N_r \tag{81}$$

where $\alpha_{k,m}$ denotes the "amplitude" of the target return and $\widetilde{\rho}_{k,m}$ is associated with the clutter, both corresponding to the k^{th} receiver and the m^{th} transmitter, and $\mathbf{w}_{k,m}$ denotes the interference at the k^{th} receive antenna corresponding to the m^{th} frequency band. Making assumptions similar to those stated in Section I leads to the following optimal detector:

$$\sum_{m=1}^{N_t}\sum_{k=1}^{N_r} \frac{\lambda_{k,m}|\theta_{k,m}|^2}{1+\lambda_{k,m}} \underset{\mathrm{H}_1}{\overset{\mathrm{H}_0}{\lessgtr}} \eta' \tag{82}$$

with $\lambda_{k,m} \triangleq \sigma_{k,m}^2 \mathbf{a}_m^H \mathbf{D}_{k,m}^2 \mathbf{a}_m$, and

$$\theta_{k,m} \triangleq \frac{\mathbf{a}_m^H \mathbf{D}_{k,m}^2 \mathbf{r}_{k,m}}{\|\mathbf{a}_m^H \mathbf{D}_{k,m}\|_2}, \tag{83}$$

$$\mathbf{D}_{k,m} \triangleq \left(\sigma_{\mathrm{c},(k,m)}^2 \mathbf{a}_m \mathbf{a}_m^H + \mathbf{M}_{k,m}\right)^{-1/2}. \tag{84}$$

Herein $\sigma_{\mathrm{c},(k,m)}^2$ and $\mathbf{M}_{k,m}$ denote the variance of $\widetilde{\rho}_{k,m}$ and covariance matrix of $\mathbf{w}_{k,m}$, respectively.

It is now straightforward to verify that the code design problem for the case of multiple transmitters can be dealt with using a modified version of (25):

$$\begin{aligned} \max_{\{\mathbf{a}_m\},\{\lambda_{k,m}\}} \quad & \sum_{m=1}^{N_t}\sum_{k=1}^{N_r} f_{\mathcal{I}}(\lambda_{k,m}) + g_{\mathcal{I}}(\lambda_{k,m}) \\ \text{subject to} \quad & \lambda_{k,m} = \sigma_{k,m}^2 \mathbf{a}_m^H (\sigma_{\mathrm{c},(k,m)}^2 \mathbf{a}_m \mathbf{a}_m^H + \mathbf{M}_{k,m})^{-1} \mathbf{a}_m, \\ & \|\mathbf{a}_m\|_2^2 \leqslant e_m \quad \forall m, \end{aligned} \tag{85}$$

where e_m denotes the maximum available transmit energy for the m^{th} transmit antenna. Next observe that the above optimization problem is separable w.r.t m. Therefore, the code design procedure associated with each transmitter can be independently handled using the proposed methods in this paper.

VII. Simulation Results

In this section, we present several numerical examples to examine the performance of the proposed algorithms. In particular, we compare the system performance for coded and uncoded (employing the code vector $\mathbf{a} = \sqrt{\frac{e}{N}}\mathbf{1}$) scenarios. Comparisons between the computational costs of Sv-MaMi and Re-MaMi are also included. Moreover, the behavior of the information-theoretic criteria is assessed when e varies.

Throughout this section, we assume the code length $N = 10$, the number of receivers $N_r = 4$, variances of the target components given by $\sigma_k^2 = 1$ (for $1 \leq k \leq 4$), and variances of the clutter components given by $(\sigma_{\mathrm{c},1}^2, \sigma_{\mathrm{c},2}^2, \sigma_{\mathrm{c},3}^2, \sigma_{\mathrm{c},4}^2) = (0.125, 0.25, .5, 1)$. Furthermore, we assume that the k^{th} interference covariance matrix $\mathbf{M}_k$ is given by $[\mathbf{M}_k]_{m,n} = (1-0.15k)^{|m-n|}$. The ROC is used to evaluate the detection performance of the system. Particularly, P_d and P_{fa} are calculated using their analytical expressions (see eqs. (32)-(34) in [14]). Then the ROC is plotted by numerically eliminating the detection threshold. The CVX toolbox [56] is used for solving the MaMi convex optimization problems.

Fig. 1(a)-(d) show the ROCs associated with the coded system (employing the optimized codes) as well as the uncoded system for $e = 10$ and $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}, \mathcal{J}, \mathcal{M}\}$. The plotted ROCs correspond to the obtained codes using Sv-MaMi, Re-MaMi with either randomization (with $L = 50$) or LS, and the uncoded system. Theses figures also show the results of PAR-constrained code design with PAR$= 1$ (i.e. constant modulus) for Sv-MaMi and Re-MaMi (LS). It can be observed that the performance of the coded system (for all $\mathcal{I}$) outperforms that of the uncoded system significantly. Furthermore, the codes obtained by Sv-MaMi lead to slightly better performance compared to the codes provided by Re-MaMi as Sv-MaMi circumvents the synthesis loss. Note also the superiority of synthesis via randomization when compared to the LS synthesis. As to the constrained design, it can be seen that imposing the PAR constraint leads to a minor performance degradation (for all criteria) when compared to the unconstrained design. The fact that Sv-MaMi (PAR$= 1$) outperforms Re-MaMi (PAR$= 1$) complies with the related observation for the unconstrained case. In this example, the detection performances corresponding to various criteria are similar. However, this behavior does not generally hold true (see e.g. [21], [34], [35] and [18] for details on this aspect).

In Fig. 2 (a)-(d), the error norm has been depicted versus the iteration number for both Sv-MaMi and Re-MaMi. The error norm for Sv-MaMi and Re-MaMi is defined as $\|\mathbf{a}^{(l+1)} - \mathbf{a}^{(l)}\|_2$ and $\|\mathbf{A}^{(l+1)} - \mathbf{A}^{(l)}\|_F$, respectively. It can be observed that Re-MaMi converges much faster than Sv-MaMi. This observation can be explained by noting that in Sv-MaMi several majorizations have been applied successively. However, the complexity per iteration of Sv-MaMi is less than that of Re-MaMi because each iteration of Sv-MaMi can be handled efficiently by solving a convex QCQP. Another observation is that for the metrics $\mathcal{B}$ and $\mathcal{D}$, both algorithms require more iterations for convergence when compared to $\mathcal{J}$ and $\mathcal{M}$. This might be due to the more complicated form of the objective functions associated with $\mathcal{B}$ and $\mathcal{D}$. Note that for $\mathcal{I} = \mathcal{M}$, Re-MaMi not only needs just one iteration to converge but also it provides the globally optimal solution to the relaxed optimization problem owing to its convexity (see (72)).

The required computation time of Sv-MaMi and Re-MaMi (employing randomization with $L = 50$)

for various criteria is shown in Table III. Due to the fact that the convergence time is dependent on the starting point as well as the stop criterion, the reported times are averaged for 100 random starting points on a standard PC (with Intel Core i5 2.8GHz CPU and 8GB memory) assuming $\xi = 10^{-4}$. It can be observed from this table that for $\mathcal{I} \in \{\mathcal{J}, \mathcal{M}\}$, Re-MaMi converges much faster than it does in the case of $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}\}$. This can be explained by considering the required iteration numbers for different $\mathcal{I}$. Furthermore, the computational times of Sv-MaMi are almost the same for all criteria. In sum, from a computational point of view, it can be concluded that Re-MaMi is preferable for $\mathcal{I} \in \{\mathcal{J}, \mathcal{M}\}$ whereas Sv-MaMi is more suitable for $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}\}$. It is also practically observed that there is no considerable difference between computational time of Re-MaMi with either LS or randomization with $L = 50$.

The behavior of various information-theoretic criteria versus the transmit energy e is investigated in Fig. 3 (a)-(d) for the coded system (using Sv-MaMi with PAR$= 1$, and without PAR constraint) as well as the uncoded system. This figure also illustrates the saturation phenomenon. We observe from Fig. 3 that a saturation of the coded system always occurs before that in the uncoded system, which was expected: employing an optimized code enables the system to perform closer to the best possible performance at lower values of e. For all criteria, an approximate decrease of 14.5 dB in the required transmit energy of the coded system (with PAR$= 1$) is observed for $e = 10$ as for Fig. 1 (see above).

TABLE III

COMPARISON OF THE AVERAGE COMPUTATIONAL TIMES (IN SEC.) OF SV-MAMI AND RE-MAMI ON A STANDARD PC

Criterion / Algorithm	$\mathcal{B}$	$\mathcal{D}$	$\mathcal{J}$	$\mathcal{M}$
Sv-MaMi	20.31	20.86	16.88	18.91
Re-MaMi	49.01	27.71	2.51	3.16

VIII. CONCLUSIONS

Multi-static radar code design schemes based on information-theoretic criteria were considered in the presence of clutter. Two general methods were proposed to tackle the highly non-linear and non-convex design optimization problems using the Majorization-Minimization (MaMi) technique. The main results can be summarized as follows:

- A discrete-time formulation of the problem as well as the associated optimal detector were presented. Due to the lack of analytical ROC, information-theoretic criteria were used as design metrics, viz.

the Bhattacharyya distance, KL-divergence, J-divergence, and the Mutual information.Using these metrics, optimization problems corresponding to the original code design problem were derived.

- A unified framework was proposed to describe all the arising optimization problems. Two methods called Sv-MaMi and Re-MaMi (based on the MaMi technique) were devised to solve these optimization problems:
  i) Sv-MaMi uses successive (linear as well as quadratic) majorizations such that each iteration of the algorithm can be handled using a convex QCQP.
  ii) Re-MaMi consists of majorization steps, rank-one constraint relaxation, and a synthesis stage. A least-squares approach and a randomization technique were used for code synthesis.
- The proposed methods were extended to PAR-constrained code design problems and to the case of multiple transmitters (with orthogonal transmission).
- Numerical examples were provided to examine the proposed methods. It was observed that Re-MaMi is computationally more efficient for $\mathcal{I} \in \{\mathcal{J}, \mathcal{M}\}$. On the other hand, for $\mathcal{I} \in \{\mathcal{B}, \mathcal{D}\}$ Sv-MaMi is preferable. The metric's saturation phenomenon, as the transmit energy increases, was also investigated.

Note that stationary targets were considered in this work. Optimal code design using information-theoretic criteria in the case of moving targets can be an interesting topic for future research.

Appendix A
Derivation of the discrete-time Model

It follows from (1) and (2) that the n^{th} sample of the output of the matched filter at the k^{th} receiver can be written as

$$\begin{aligned} r_{k,n} = r_k(t) \star \phi^*(-t)\Big|_{t=(n-1)T_p+\tau_k} &= \int_{-\infty}^{+\infty} r_k(\tau)\phi^*(\tau - [n-1]T_p - \tau_k)d\tau \\ &= \int_{-\infty}^{+\infty} \alpha_k \sum_{i=1}^{N} a_i \phi(\tau - [i-1]T_p - \tau_k)\phi^*(\tau - [n-1]T_p - \tau_k)d\tau \\ &+ \int_{-\infty}^{+\infty} \sum_{v=1}^{N_c} \rho_{k,v} \sum_{i=1}^{N} a_i \phi(\tau - [i-1]T_p - \tau_{k,v})\phi^*(\tau - [n-1]T_p - \tau_k)d\tau \\ &+ \int_{-\infty}^{+\infty} w_k(\tau)\phi^*(\tau - [n-1]T_p - \tau_k)d\tau. \end{aligned} \tag{86}$$

Let Q_u denote the u^{th} integral in the right-hand side of the above equation. Since $\{\phi(t-[n-1]T_p)\}_{n=1}^{N}$ are non-overlapping and have unit energy, Q_1 can be simplified as

$$Q_1 = \alpha_k \sum_{i=1}^{N} a_i \delta[i-n] = a_n \alpha_k. \tag{87}$$

Furthermore, we have that

$$Q_2 = \sum_{v=1}^{N_c} \rho_{k,v} \left(\sum_{i=1}^{N} a_i \Psi_{n,i}(\tau_k - \tau_{k,v}) \right) \tag{88}$$

where $\Psi_{n,i}(t)$ is the cross-correlation function of $\phi(\tau-[i-1]T_p)$ and $\phi(\tau-[n-1]T_p)$ defined by

$$\Psi_{n,i}(t) \triangleq \int_{-\infty}^{+\infty} \phi(\tau-[i-1]T_p - t)\phi^*(\tau-[n-1]T_p) d\tau. \tag{89}$$

For unambiguous-range clutter scatterers (i.e. scatterers with $\tau_{k,v} \leq T_p$) [24], $\Psi_{n,i}(\tau_k - \tau_{k,v})$ is zero for $i \neq n$ because $\phi(t-[i-1]T_p - \tau_{k,v})$ and $\phi(t-[n-1]T_p-\tau_k)$ are non-overlapping[2]. Therefore, Q_2 can be rewritten as

$$Q_2 = a_n \left(\sum_{v=1}^{N_c} \rho_{k,v} \Psi_{n,n}(\tau_k - \tau_{k,v}) \right) \triangleq a_n \widetilde{\rho}_k. \tag{90}$$

Note that $w_{k,n} \triangleq Q_3$ represents the filtered version of the interference. Finally, we can simplify (86) as

$$r_{k,n} = a_n \alpha_k + a_n \widetilde{\rho}_k + w_{k,n}, \quad \text{for} \quad k = 1,2,..,N_r \quad \text{and} \quad n = 1,2,..,N$$

According to Assumption 3, $\beta_{k,v} \triangleq \rho_{k,v}\Psi_{n,n}(\tau_k - \tau_{k,v})$ are independent RVs, for $v = 1,2,...,N_c$. Consequently, $\widetilde{\rho}_k = \sum_{v=1}^{N_c} \beta_{k,v}$ can be modeled, using the central limit theorem [25], as a zero-mean complex Gaussian RV with variance $\sigma^2_{c,k}$. Note that $\sigma^2_{c,k}$ can be calculated using $\Psi_{n,n}(\cdot)$ and the distribution of the $(\tau_k - \tau_{k,v})$ [2].

APPENDIX B

MONOTONICALLY INCREASING BEHAVIOR OF THE FUNCTION $q_{\mathcal{I}}(\lambda_k)$

For $\mathcal{I} = \mathcal{B}$, we have $q_{\mathcal{I}}(\lambda_k) = \log \frac{1+.5\lambda_k}{\sqrt{1+\lambda_k}}$. Therefore, the first-order derivative of $q_{\mathcal{B}}(\lambda_k)$ is given by

$$\frac{d}{d\lambda_k} q_{\mathcal{B}}(\lambda_k) = \frac{0.25\lambda_k}{(1+\lambda_k)(1+0.5\lambda_k)}. \tag{91}$$

Similarly, for the first-order derivative of $q_{\mathcal{D}}(\lambda_k)$ we have

$$\frac{d}{d\lambda_k} q_{\mathcal{D}}(\lambda_k) = \frac{\lambda_k}{(\lambda_k+1)^2}. \tag{92}$$

[2]Note that $\tau_k \geq \tau_p$, otherwise τ_k corresponds to a blind range of the system [24].

As to the J-divergence, one can easily verify that

$$\frac{d}{d\lambda_k} q_{\mathcal{J}}(\lambda_k) = \frac{\lambda_k^2 + 2\lambda_k}{(1+\lambda_k)^2}. \tag{93}$$

Due to the fact that the right-hand side in eqs. (91), (92), and (93) are non-negative for $\lambda_k \geq 0$, the function $q_{\mathcal{I}}(\lambda_k)$ is monotonically increasing for $\mathcal{I}$ in $\{\mathcal{B}, \mathcal{D}, \mathcal{J}\}$. Moreover, the case of $\mathcal{I} = \mathcal{M}$ simply leads to the monotonically increasing function $q_{\mathcal{M}}(\lambda_k) = \log(1 + \lambda_k)$.

APPENDIX C
PROOF OF LEMMA 4

First note that $\widetilde{h}(\widetilde{\mathbf{A}}) = h(\widetilde{\mathbf{A}})$. In addition, $\widetilde{h}(\mathbf{A}) \leq \mathrm{tr}\{(\mathbf{M}_k + \beta_k \mathbf{A})^{-1}\mathbf{M}_k\}$ for every pair of positive semidefinite matrices $\mathbf{A}, \widetilde{\mathbf{A}} \in \mathbb{C}^{N\times N}$ if

$$(\mathbf{M}_k + \beta_k \mathbf{A})^{-1} - (\mathbf{M}_k + \beta_k \widetilde{\mathbf{A}})^{-1} + (\mathbf{M}_k + \beta_k \widetilde{\mathbf{A}})^{-1}(\beta_k \mathbf{A} - \beta_k \widetilde{\mathbf{A}})(\mathbf{M}_k + \beta_k \widetilde{\mathbf{A}})^{-1} \succeq \mathbf{0}. \tag{94}$$

Observe that $(\beta_k \mathbf{A} - \beta_k \widetilde{\mathbf{A}}) = (\mathbf{M}_k + \beta_k \mathbf{A}) - (\mathbf{M}_k + \beta_k \widetilde{\mathbf{A}})$. Therefore, using the variables $\mathbf{X} = (\mathbf{M}_k + \beta_k \mathbf{A})$ and $\mathbf{Y} = (\mathbf{M}_k + \beta_k \widetilde{\mathbf{A}})$, one can rewrite the left-hand side of (94) as

$$\mathbf{X}^{-1} - \mathbf{Y}^{-1} + \mathbf{Y}^{-1}(\mathbf{X} - \mathbf{Y})\mathbf{Y}^{-1} = (\mathbf{I} - \mathbf{Y}^{-1}\mathbf{X})\mathbf{X}^{-1}(\mathbf{I} - \mathbf{X}\mathbf{Y}^{-1}).$$

Now it is straightforward to verify that the right-hand side of the above equation is always positive semi-definite as $\mathbf{X}^{-1} \succ \mathbf{0}$ and $(\mathbf{I} - \mathbf{X}\mathbf{Y}^{-1}) = (\mathbf{I} - \mathbf{Y}^{-1}\mathbf{X})^H$.

REFERENCES

[1] W. D. Rummler, "A technique for improving the clutter performance of coherent pulse trains," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 3, pp. 898–906, Nov. 1967.

[2] D. F. Delong and E. M. Hofstteter, "On the design of optimum radar waveforms for clutter rejection," *IEEE Trans. Inf. Theory*, vol. 13, pp. 454–463, Jul. 1967.

[3] ——, "The design of clutter-resistant radar waveforms with limited dynamic range," *IEEE Trans. Inf. Theory*, vol. 15, pp. 376–385, May 1969.

[4] L. J. Spafford, "Optimum radar signal processing in clutter," *IEEE Trans. Inf. Theory*, vol. 14, pp. 734–743, Sep. 1968.

[5] A. V. Balakrishnan, "Signal design for a class of clutter channels," *IEEE Trans. Inf. Theory*, pp. 170–173, Jan. 1968.

[6] F. R. Castella and J. R. Moore, "Improved target detection in littoral environment," in *Radar 97*, Oct. 1997, pp. 429–433.

[7] S. U. Pillai, H. S. Oh, D. C. Youla, and J. R. Guerci, "Optimum transmit-receiver design in the presence of signal-dependent interference and channel noise," *IEEE Trans. Inf. Theory*, vol. 46, pp. 577–584, May 2000.

[8] S. M. Kay, "Optimal signal design for detection of Gaussian point targets in stationary Gaussian clutter/reverberation," *IEEE J. Sel. Topics Signal Process.*, vol. 1, pp. 31–41, Jun. 2007.

[9] P. Stoica, H. He, and J. Li, "Optimization of the receive filter and transmit sequence for active sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1730 –1740, Apr. 2012.

[10] A. Aubry, A. D. Maio, A. Farina, and M. Wicks, “Knowledge-aided (potentially cognitive) transmit signal and receive filter design in signal-dependent clutter,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, pp. 93–117, Jan. 2013.

[11] M. R. Bell, “Information theory and radar waveform design,” *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1578–1597, 1993.

[12] R. A. Romero, J. Bae, and N. A. Goodman, “Theory and application of SNR and mutual information matched illumination waveforms,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, pp. 912–927, Apr. 2011.

[13] S. M. Sowelam and A. H. Tewfik, “Waveform selection in radar target classification,” *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 1014–1029, 2000.

[14] M. M. Naghsh and M. Modarres-Hashemi, “Exact theoretical performance analysis of optimum detector for statistical MIMO radars,” *IET Journal on Radar, Sonar, and Navigation*, vol. 6, pp. 99–111, 2012.

[15] S. M. Kay, “Waveform design for multistatic radar detection,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, pp. 1153–1165, Jul. 2009.

[16] A. De Maio, M. Lops, and L. Venturino, “Diversity integration trade-off in MIMO detection,” *IEEE Trans. Signal Process.*, vol. 56, pp. 5051–5061, Oct. 2008.

[17] P. Stoica, J. Li, and Y. Xie, “On probing signal design for MIMO radar,” *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4151–4161, 2007.

[18] E. Grossi and M. Lops, “Space-time code design for MIMO detection based on Kullback-Leibler divergence,” *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3989–4004, 2012.

[19] B. Tang, J. Tang, and Y. Peng, “MIMO radar waveform design in colored noise based on information theory,” *IEEE Trans. Signal Process.*, vol. 58, pp. 4684–4697, Sep. 2010.

[20] R. S. Blum and Y. Yang, “MIMO radar waveform design based on mutual information and minimum mean-square error estimation,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, pp. 330–343, Jan. 2007.

[21] T. Kailath, “The divergence and Bhattacharyya distance measures in signal selection,” *IEEE Trans. Commun.*, vol. 15, pp. 52–60, Feb. 1967.

[22] X. Song, P. Willett, S. Zhou, and P. Luh, “The MIMO radar and jammer games,” *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 687–699, Feb. 2012.

[23] S. Sen and A. Nehorai, “OFDM MIMO radar with mutual-information waveform design for low-grazing angle tracking,” *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3152–3162, 2010.

[24] M. Skolnik, *Radar Handbook*, 3rd ed. New York: McGraw-Hill, 2008.

[25] R. Manasse, “The use of pulse coding to discriminate against clutter,” *Technical report, MIT Lincoln Lab*, 1957.

[26] J. Li and P. Stoica, Eds., *MIMO Radar Signal Processing*. New York: Wiley, 2008.

[27] A. Sheikhi and A. Zamani, “Temporal coherent adaptive target detection for multi-input multi-output radars in clutter,” *Radar, Sonar & Navigation, IET*, vol. 2, no. 2, pp. 86–96, 2008.

[28] F. Gini and M. Rangaswamy, *Knowledge based radar detection, tracking and classification*. Wiley-Interscience, 2008.

[29] J. R. Gureci, *Cognitive Radar: The Knowledge-aided Fully Adaptive Approach*. Artech House, 2010.

[30] S. Haykin, “Cognitive radars,” *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 30–40, Jan. 2006.

[31] A. Aubry, A. De Maio, M. Piezzo, A. Farina, and M. Wicks, “Cognitive design of the receive filter and transmitted phase code in reverberating environment,” *Radar, Sonar Navigation, IET*, vol. 6, no. 9, pp. 822–833, 2012.

[32] S. M. Kay, *Fundamentals of Statistical Signal Processing-Volume II: Detection Theory*, 1st ed. New Jersey: Prentice Hall, 1998.

[33] A. Jain, P. Moulin, M. I. Miller, and K. Ramchandran, "Information-theoretic bounds on target recognition performance based on degraded image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 1153–1166, Sep. 2002.

[34] S. Kullback, *Information theory and statistics*. New York: John Wiley and Sons, 1968.

[35] H. Kobayashi and J. B. Thomas, "Distance measures and related criteria," in *Proc. 5th Annu. Allerton Conf. Circuit and System Theory*, 1967, pp. 491–500.

[36] T. Kadota and L. Shepp, "On the best finite set of linear observables for discriminating two gaussian signals," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 278–284, 1967.

[37] T. Grettenberg, "Signal selection in communication and radar systems," *IEEE Trans. Inf. Theory*, vol. 9, no. 4, pp. 265–275, 1963.

[38] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[39] F. Gini, A. De Maio, and L. Patton, *Waveform design and diversity for advanced radar systems*. Institution of engineering and technology, 2012, vol. 22.

[40] F. Kanaya and K. Nakagawa, "On the practical implication of mutual information for statistical decision making," *IEEE Trans. Inf. Theory*, vol. 37, no. 4, pp. 1151–1156, 1991.

[41] S. D. Briles, "Information-theoretic performance bounding of Bayesian identifiers," in *Optical Engineering and Photonics in Aerospace Sensing*. International Society for Optics and Photonics, 1993, pp. 255–266.

[42] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2005.

[43] P. Stoica and Y. Selen, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 112 – 114, Jan. 2004.

[44] D. R. Hunter and K. Lange, "A toturial on MM algorithm," *The American Statistician*, vol. 58, pp. 30–37, February 2004.

[45] K. Lange, *Numerical Analysis for Statisticians*. New York: New York.: Springer-Verlag, 1999.

[46] J. D. Leeuw and K. Lange, "Sharp quadratic majorization in one dimension," *Computational Statistics and Data Analysis*, vol. 53, pp. 2471–2484, May 2009.

[47] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2407 – 2423, Sep. 2004.

[48] T. Naghibi and F. Behnia, "MIMO radar waveform design in the presence of clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, pp. 770–781, Apr. 2011.

[49] 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, May 2010.

[50] G. Pataki, "On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues," *Mathematics of operations research*, vol. 23, pp. 339–358, 1998.

[51] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge University Press, 2004.

[52] 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, pp. 2683–2697, Jun. 2011.

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

[54] 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.

[55] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, “Designing structured tight frames via an alternating projection method,” *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–208, Jan. 2005.

[56] M. Grant and S. Boyd. (2012, February) CVX package. [Online]. Available: http://www.cvxr.com/cvx

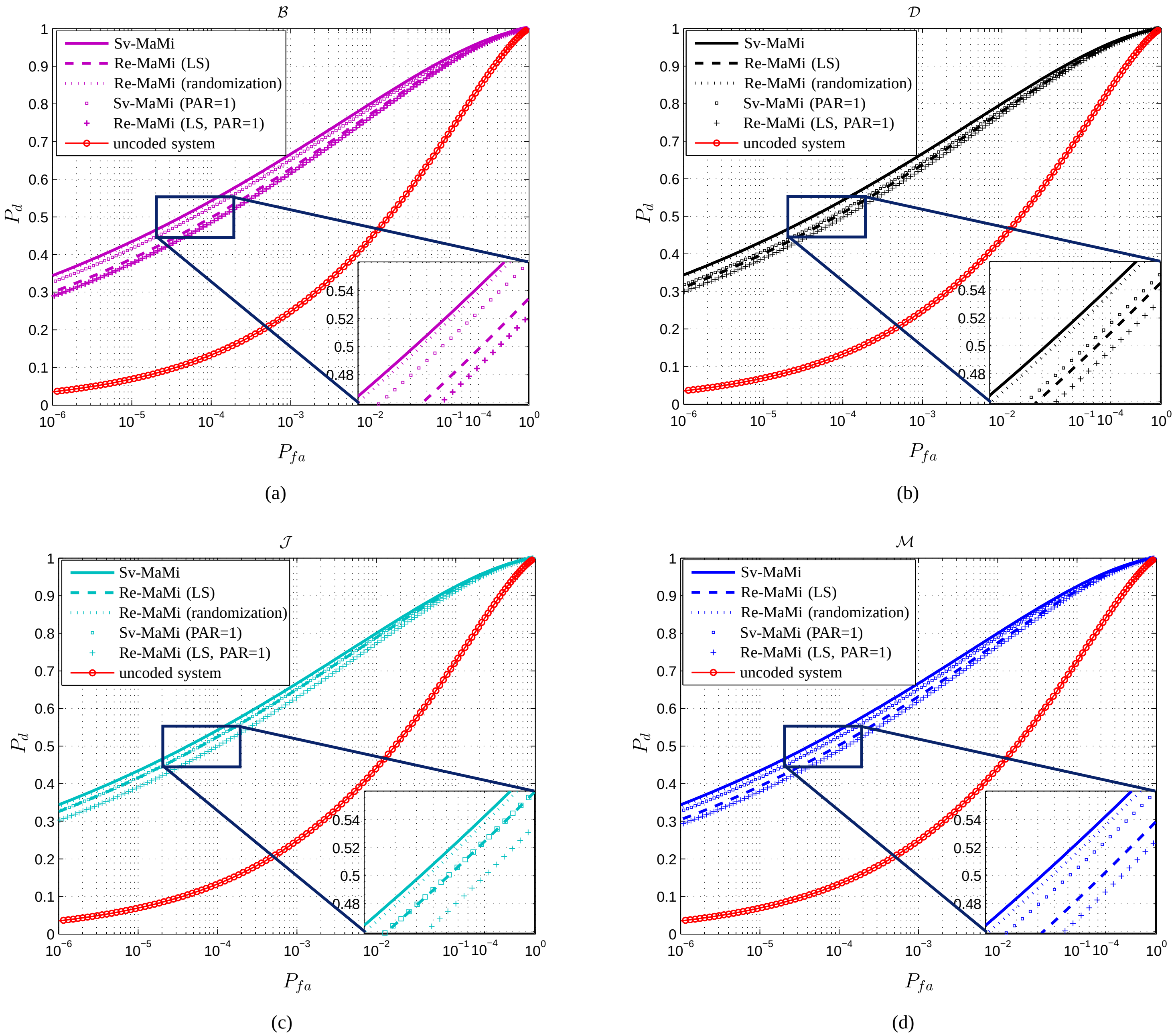

Fig. 1. ROCs corresponding to the obtained codes using Sv-MaMi and Re-MaMi (both PAR-constrained and unconstrained) as well as the uncoded system for different design metrics: a) $\mathcal{B}$, b) $\mathcal{D}$, c) $\mathcal{J}$, and d) $\mathcal{M}$. For unconstrained design using Re-MaMi algorithm, results of both LS synthesis and randomization are shown.

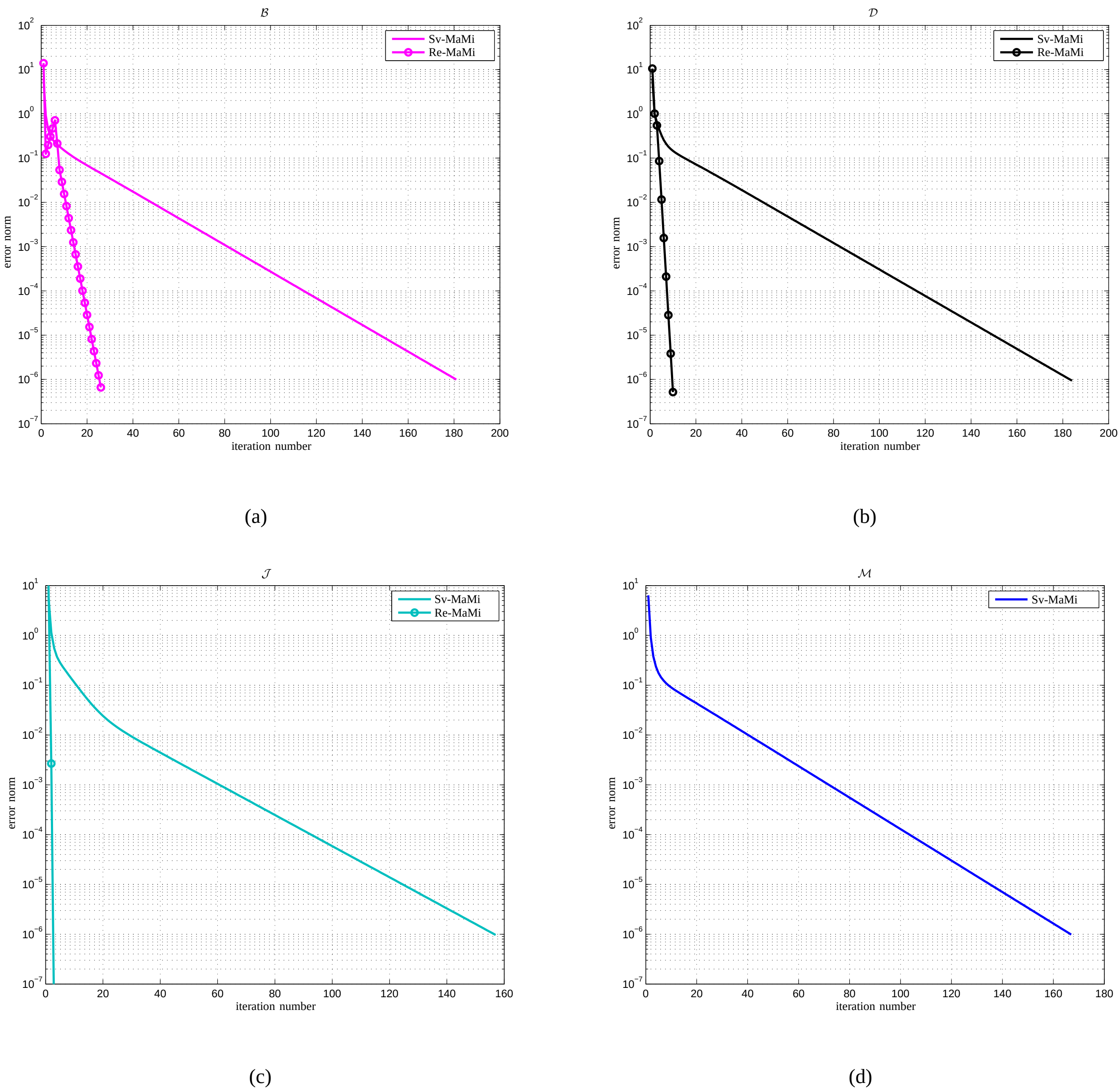

Fig. 2. Error norm versus iteration number for Sv-MaMi/Re-MaMi and different design metrics: a) $\mathcal{B}$, b) $\mathcal{D}$, c) $\mathcal{J}$, and d) $\mathcal{M}$. For the case of $\mathcal{M}$, Re-MaMi converges in one iteration.

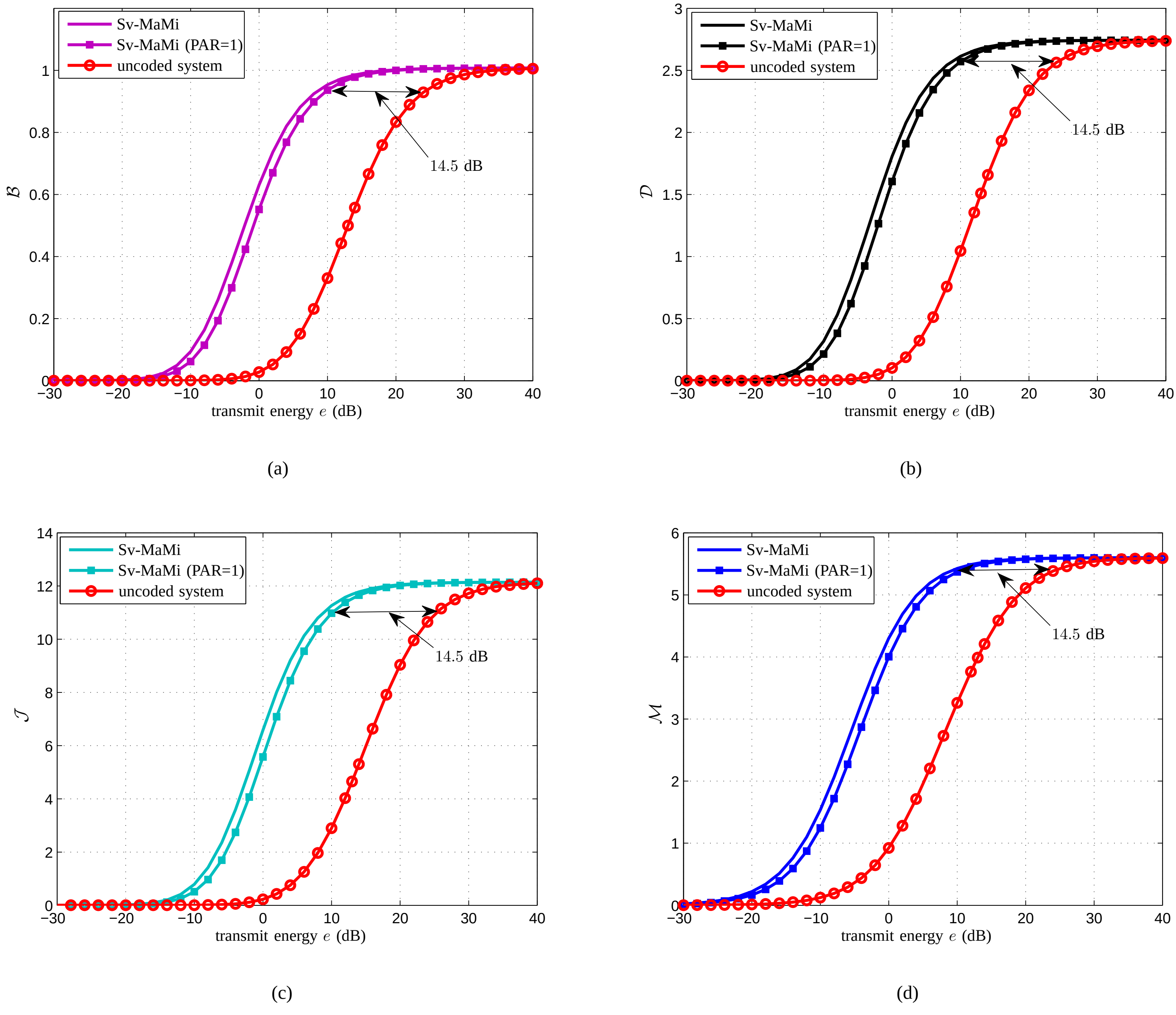

Fig. 3. Behavior of various information-theoretic criteria versus transmit energy e for the coded and the uncoded systems: a) $\mathcal{B}$, b) $\mathcal{D}$, c) $\mathcal{J}$, and d) $\mathcal{M}$. Results for Sv-MaMi with PAR$=1$ and with no PAR constraint are shown.