Relaxed Biquadratic Optimization for Joint Filter-Signal Design in Signal-Dependent STAP

Sean M. O'Rourke, *Member, IEEE*, Pawan Setlur, *Member, IEEE*, Muralidhar Rangaswamy, *Fellow, IEEE*, and A. Lee Swindlehurst, *Fellow, IEEE*

***Abstract*—We investigate an alternative solution method to the joint signal-beamformer optimization problem considered in a recent publication devoted to this topic. First, we directly demonstrate that the problem, which minimizes the received noise, interference, and clutter power under a minimum variance distortionless response constraint, is generally nonconvex and we provide insight into the nature of the nonconvexity. Second, we employ the theory of biquadratic optimization and semidefinite relaxations to produce a relaxed version of the problem, which we show to be *convex*. The optimality conditions of this relaxed problem are examined and a variety of potential solutions are found, both analytically and numerically. These solutions are then compared to existing alternating minimization schemes.**

***Index Terms*—Capon beamformer, convex optimization, joint design, semidefinite relaxation, space time adaptive radar, waveform design.**

I. Introduction

In 1972, at a NATO conference in the East Midlands of England, two luminaries of signal processing agreed that an adaptive system was incapable of being simultaneously spatially and temporally optimal. Agreeing with M. Mermoz's proposition, the Naval Underwater System Center's Norman Owsley asserted that since "the spectrum of the signal must be known *a priori*" for temporal optimality, "the post beamformer temporal processor cannot be fully adaptive and fully optimum (sic) simultaneously" [1]. Owing to the technology of the day, Owsley's assertion refers primarily to adaptivity and optimality-on-receive; however, this still leaves open the question of what spectrum to transmit – something P.M. Woodward observed "remains substantially unanswered" many years before [2]. Given the modern flexibility of waveform generation, it is worthwhile to ask if we can find simultaneously optimal transmit waveforms and receive filters over a given epoch for a known or estimated scattering environment. While this does not necessarily violate the optimality/adaptivity tradeoff above, it does present the potential to reframe the optimization problem beyond the constraints of optimality-on-transmit and optimality-on-receive. In this paper, we present a first step in that direction for radar space-time adaptive processing (STAP), which in turn is a significant expansion of the work of [3].

In the general scenario, we assume that an airborne radar equipped with an array of antenna elements observes a moving target below. Furthermore, this radar can change its transmitted waveform every coherent processing interval (CPI), instead of on a per-pulse basis. In order to develop a strategy for waveform design, we consider a STAP model that includes fast-time (range) samples as well as the usual slow-time and spatial samples. This is a departure from traditional STAP, which operates on spatio-Doppler responses from the radar after matched filtering [4], [5], but recent advances in the literature have considered incorporating fast-time data for more accurate clutter modeling, both transmitter- and jammer-induced [6], [7].

An obvious result of including the fast-time data in the model is that the clutter representation is now signal-dependent, since in airborne STAP, the dominant clutter source is non-target ground reflections that persist over all range bins. We assume that, if modeled as a random process, the clutter is uncorrelated with any interference or noise (which, unlike the clutter, are assumed to have no signal dependence), and the related clutter correlation matrix is also signal dependent. If we formulate our waveform-filter design in the typical minimum variance distortionless-response (MVDR) framework [8], this dependence in the correlation structure leads to what researchers have assumed to be a non-convex optimization problem [9]–[12]. However, the authors in [3] correctly identified that for a fixed transmit waveform, the problem is convex in the receive filter, and vice versa. This led to a collection of design algorithms based on alternating minimization, a reasonable heuristic. While the alternating minimization method is useful, it has no claims of optimality, nor do other methods in the literature that dealt with signal-dependent interference in other contexts, like the single sensor radar and reverberant channel of [13]. Furthermore, none of this work directly demonstrated the non-convexity of the problem. We will ameliorate some of these concerns in this paper, first by directly proving the non-convexity of the objective, then relating it to the existing literature on

Manuscript received May 31, 2017; revised October 3, 2017; accepted November 3, 2017. Date of publication November 20, 2017; date of current version January 26, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rui Zhang. This work was supported by the Air Force Office of Scientific Research through Project 17RYCOR481 under the Dynamic Data Driven Application Systems Program. *(Corresponding author: Sean M. O'Rourke.)*

S. M. O'Rourke is with the Sensors Directorate, US Air Force Research Laboratory, Wright-Patterson Air Force Base, OH 45433 USA, and also with the University of California, Irvine, Irvine CA 92697-2625, USA (e-mail: sean.orourke.3@us.af.mil).

P. Setlur and M. Rangaswamy are with the Sensors Directorate, US Air Force Research Laboratory, Wright-Patterson Air Force Base, OH 45433 USA (e-mail: pawan.setlur.1@us.af.mil; muralidhar.rangaswamy@us.af.mil).

A. L. Swindlehurst is with the University of California, Irvine, Irvine CA 92697-2625, USA (e-mail: swindle@uci.edu).

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

Digital Object Identifier 10.1109/TSP.2017.2775592

biquadratic programming (BQP). While the biquadratic program is demonstrably non-convex, it is possible to relax it to a convex quadratic program using semidefinite programming (SDP) (see [14] for details on SDP and [15], [16] on relaxing the BQP). This relaxation will permit us to efficiently solve the problem computationally and analytically up to a matrix completion, as well as reveal important structural information about the solution that matches with engineering intuition.

Before continuing, we outline some mathematical notation that will be used throughout the paper. The symbol $\mathbf{x} \in \mathbb{R}^{N\times 1}(\mathbf{x} \in \mathbb{C}^{N\times 1})$ indicates a column vector of real (complex) values, while $\mathbf{X} \in \mathbb{R}^{M\times N}(\mathbf{X} \in \mathbb{C}^{M\times N})$ indicates an N row, M column matrix of real (complex) values. The sets $\mathbb{H}_n, \mathbb{H}_n^+$, and $\mathbb{H}_n^{++}$ represent Hermitian, positive semidefinite, and positive definite complex matrices, respectively. The latter two sets induce the orderings $\succeq$ and $\succ$. The superscripts T, $*$, and H indicate the transpose, conjugate, and Hermitian (conjugate) transpose of a matrix or vector. The symbol $\otimes$ indicates a Kronecker product. The operator vec turns a matrix into a vector with the matrix stacked columnwise – that is, for $\mathbf{A} \in \mathbb{R}^{p\times q}, \mathrm{vec}(\mathbf{A}) \in \mathbb{R}^{pq\times 1}$. Special matrices that will recur often include the $n \times n$ identity matrix $\mathbf{I}_n$, the $n \times 1$ vector of all ones $\mathbf{1}_n$, the $n \times m$ all-zero matrix $\mathbf{0}_{n\times m}$ and, via Magnus & Neudecker [17], the (p, q) *commutation matrix* $\mathbf{K}_{p,q} \in \mathbb{R}^{pq\times pq}$ that is uniquely determined by the relation $\mathrm{vec}(\mathbf{A}^T) = \mathbf{K}_{p,q}\,\mathrm{vec}(\mathbf{A})$ for $\mathbf{A}$ as above. We will use calligraphic letters like $\mathcal{C}$ to indicate tensors or multilinear operators (like the Hessian matrix). Additionally, the vectorizations of various matrices will be indicated by the equivalent lowercase Greek symbol in bold – for example, a matrix $\mathbf{B}$ will have the vectorization $\boldsymbol{\beta} = \mathrm{vec}(\mathbf{B})$.

The rest of the paper continues as follows: First, we will outline the general signal model and relate it to the work on transfer functions by [18]. Section III describes the joint design problem and demonstrates that the problem is non-convex unless all clutter patches are known to be nulled a priori. Section IV introduces the biquadratic problem and methods to relax the joint design into a convex quadratic semidefinite program. We then attempt to find analytic solutions and insights into this relaxed problem in Section V and present proofs of certain structural aspects of any solution in Section VI. These solutions are generally verified in Section VII through simulation, demonstrating the resulting rank-one solution can outperform multiple alternating minimization techniques. Finally, we summarize our conclusions and highlight future research paths in Section VIII.

II. STAP Model

Consider a radar that consists of a calibrated airborne linear array of M identical sensor elements, where the first element is the phase center and also acts as the transmitter. During the transmission period, the radar probes the environment with a number of pulses $s(t)$ of width T seconds and bandwidth B Hz at a carrier frequency f_o. Within each burst, we transmit L pulses at a rate of f_p (i.e. the pulses are transmitted every $T_p = 1/f_p$ seconds) and collect them in a coherent processing interval (CPI). We also assume that the phase center is located at $\mathbf{x}_r$ and the platform moves at a rate that is constant over the CPI.

Assume the probed environment contains a target that lies at an azimuth θ_t and an elevation ϕ_t relative to the array phase center, moving with a relative velocity vector $\boldsymbol{\delta}v = \left[\delta v_x\ \delta v_y\ \delta v_z\right]^T$. If we assume that the array interelement spacing d is small relative to the distance between the platform and the target, then the target's doppler shift is independent of the element index and given by

$$f_d = 2f_o \frac{\boldsymbol{\delta}v^T\left[\sin(\phi_t)\sin(\theta_t)\ \ \sin(\phi_t)\cos(\theta_t)\ \ \cos(\phi_t)\right]}{c} \tag{1}$$

where c is the speed of light.

Let us now assume that we discretely sample the pulse $s(t)$ into N samples, resulting in the sample vector $\mathbf{s} = [s_1, s_2, \cdots, s_N]^T \in \mathbb{C}^N$. Assuming the data is aligned to a common reference and given other assumptions from [19], we can say that at the target range gate τ_t, the combined target response over the entire CPI can be represented by a vector $\mathbf{y}_t \in \mathbb{C}^{NML}$ given by

$$\mathbf{y_t} = \rho_t \mathbf{v}_t(f_d) \otimes \mathbf{s} \otimes \mathbf{a}_t(\theta_t, \phi_t) \tag{2}$$

where ρ_t is the complex backscattering coefficient from the target, the vector $\mathbf{v}_t(f_d) \in \mathbb{C}^L$ is the doppler steering vector whose ith element is given by $e^{-\jmath 2\pi f_d (i-1)T_p}$, and the vector $\mathbf{a}_t(\theta_t, \phi_t) \in \mathbb{C}^M$ is the spatial steering vector whose ith element is given by $e^{-\jmath 2\pi(i-1)\vartheta}$ where $\vartheta = d\sin(\theta_t)\sin(\phi_t)f_o/c$. The presence of fast-time samples $\mathbf{s}$ differentiates the model from the standard STAP scenario.

The target return $\mathbf{y}_t$ is corrupted by a variety of undesired returns from the environment – noise, signal-independent interference, and clutter:

$$\tilde{\mathbf{y}} = \mathbf{y_t} + \mathbf{y_n} + \mathbf{y_i} + \mathbf{y_c} = \mathbf{y_t} + \mathbf{y_u}, \tag{3}$$

where the subscripts n, i, c, and u stand for noise, interference, clutter and undesired, respectively. We assume that these undesired energy sources are statistically uncorrelated from each other. We shall subsequently describe each of these sources, starting with the noise.

We assume the noise is zero mean and identically distributed across all sensors, pulses, and fast time samples. The covariance matrix of $\mathbf{y_n}$ is denoted $\mathbf{R_n} \in \mathbb{C}^{NML\times NML}$.

The interference term consists of K known interference sources that correspond to a zero-mean random process spread over all pulses and fast time samples. Assume that the kth interferer is located at the azimuth-elevation pair (θ_k, ϕ_k). In the lth PRI, we assume that the waveform is a complex continuous-time signal $\alpha_{kl}(t)$ which, when sampled, yields the vector $\boldsymbol{\alpha}_{kl} \in \mathbb{C}^N$, similar in form to $\mathbf{s}$. Stacked across all PRIs, we obtain the random vector $\boldsymbol{\alpha}_k = [\boldsymbol{\alpha}_{k0}^T\ \boldsymbol{\alpha}_{k1}^T\ \cdots\ \boldsymbol{\alpha}_{k(L-1)}^T]^T \in \mathbb{C}^{NL}$, whose covariance matrix we define as $\mathbf{E}\{\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^H\} = \mathbf{R}_{\boldsymbol{\alpha},k}$. Then, the response from the kth interferer can be modeled as

$$\mathbf{y}_{\mathbf{i},k} = \boldsymbol{\alpha}_k \otimes \mathbf{a_i}(\theta_k, \phi_k) \tag{4}$$

where, as with the target, $\mathbf{a}_i(\theta_k, \phi_k)$ is the array response to the interferer. The covariance of $\mathbf{y}_{\mathbf{i},k}$ is then $\mathbf{R}_{\boldsymbol{\alpha},k} \otimes \mathbf{a_i}(\theta_k, \phi_k)\mathbf{a}_i(\theta_k, \phi_k)^H$, since expectations follow through Kronecker products that don't have a random dependence. If we assume the interferers are uncorrelated with each other, then the

overall covariance of the combined signal-independent interference $\mathbf{y}_{\mathtt{i}}$ is

$$\mathbf{R}_{\mathtt{i}} = \sum_{k=1}^{K} \mathbf{R}_{\boldsymbol{\alpha},k} \otimes \mathbf{a}_{\mathtt{i}}(\theta_k, \phi_k)\mathbf{a}_i(\theta_k, \phi_k)^H. \tag{5}$$

In future sections, we will use the notation $\mathbf{R}_{\mathtt{ni}} = \mathbf{R}_{\mathtt{n}} + \mathbf{R}_{\mathtt{i}}$ to denote the combined noise and interference covariance matrix.

In airborne radar applications, the most significant clutter source is the ground, which produces returns persistent throughout all range gates up to the horizon. Though other clutter sources exist, like large discrete objects, vegetation, and targets not currently being surveilled, the specific stochastic model we apply here only concerns ground clutter. However, we note that a later formulation in this paper *may* be amenable to considering those other sources in a manner that recalls efforts in the literature on channel estimation (see, for example, [18]). Regardless of the source, a major assumption in all works on signal-dependent interference mitigation is that there is some prior knowledge of the clutter's response to a given input signal. While this can lead to a "chicken-or-egg" scenario, as mentioned by [3], such information can be obtained in a variety of ways: physics-based scattering models, estimation from topographical information and/or knowledge-aided databases (as in DARPA's KASSPER program [20]), or occasional direct probing and estimation of the environment with "known-good" pilot signals (see, e.g., [18]). In this work, we will consider an idealized physics-based approach, but note this does not limit our method's applicability to more realistic scenarios.

For now, assume the presence of Q clutter patches, each comprising P distinct scatterers. As with the target, the return from the pth scatterer in the qth patch, located spatially at the azimuth-elevation pair (θ_{pq}, ϕ_{pq}), maintains a Kronecker structure given by

$$\gamma_{pq}\mathbf{v}(f_{\mathtt{c},pq}) \otimes \mathbf{s} \otimes \mathbf{a}(\theta_{pq}, \phi_{pq})$$

where the returned complex reflectivity is γ_{pq} and the Doppler shift observed by the platform is $f_{\mathtt{c},pq}$. This Doppler shift is solely induced by the platform motion, characterized by the aforementioned velocity vector $\dot{\mathbf{x}}_{\mathtt{r}}$, and is given by

$$f_{\mathtt{c},pq} = 2f_o \frac{\dot{\mathbf{x}}_{\mathtt{r}}^T \left[\sin(\phi_{pq})\sin(\theta_{pq}) \;\; \sin(\phi_{pq})\cos(\theta_{pq}) \;\; \cos(\phi_{pq})\right]}{c}. \tag{6}$$

Thus, the overall response from the qth clutter patch is

$$\mathbf{y}_{\mathtt{c},q} = \sum_{p=1}^{P} \gamma_{pq}\mathbf{v}(f_{\mathtt{c},pq}) \otimes \mathbf{s} \otimes \mathbf{a}(\theta_{pq}, \phi_{pq}). \tag{7}$$

In order to define the covariance matrix of the clutter, we define the qth combining matrix $\mathbf{B}_q \in \mathbb{C}^{NML \times P}$ as the matrix whose pth column is $\mathbf{v}(f_{\mathtt{c},pq}) \otimes \mathbf{s} \otimes \mathbf{a}(\theta_{pq}, \phi_{pq})$, $i \in \{1, \ldots, P\}$ and the covariance of the reflectivity vector $[\gamma_{1q} \;\; \gamma_{2q} \;\; \cdots \;\; \gamma_{Pq}]^T$ given by $\mathbf{R}_{\gamma}^{pq} \in \mathbb{C}^{P \times P}$. Then, the overall covariance matrix for the patch is

$$\mathbf{R}_{\boldsymbol{\gamma}}^{q} = \mathbf{B}_q \mathbf{R}_{\gamma}^{pq} \mathbf{B}_q^H \tag{8}$$

If we assume that the scatterers in one patch are uncorrelated with the scatterers in any other patch, then the total clutter response is $\mathbf{y}_{\mathtt{c}} = \sum_{q=1}^{Q} \mathbf{y}_{\mathtt{c},q}$ and its covariance is given by

$$\mathbf{R}_{\mathtt{c}} = \sum_{q=1}^{Q} \mathbf{R}_{\boldsymbol{\gamma}}^{q} \tag{9}$$

We will denote the overall undesired response covariance matrix as $\mathbf{R}_{\mathtt{u}} = \mathbf{R}_{\mathtt{ni}} + \mathbf{R}_{\mathtt{c}}(\mathbf{s})$.

Since this is a rather cumbersome model, we can simplify our description of the clutter as follows. Let us assume that our range resolution is large enough that we cannot resolve individual scatterers in each patch – as mentioned in [3], this is typical in STAP applications. Thus, we can regard each scatterer in the patch as lying within the same range gate and having approximately equal Doppler shifts, hence $f_{\mathtt{c},pq} \approx f_{\mathtt{c},q}$. Similarly, if we assume far-field operation, the scatterers will lie in approximately the same angular resolution cell centered at (θ_q, ϕ_q), which means $\theta_{pq} \approx \theta_q, \phi_{pq} \approx \phi_q$. Given this simplification, we can modify our representation of the patch response and its covariance.

Under this assumption, the per-patch clutter response is

$$\mathbf{y}_{\mathtt{c},q} = \gamma_q \mathbf{v}(f_{\mathtt{c},q}) \otimes \mathbf{s} \otimes \mathbf{a}(\theta_q, \phi_q)$$

where $\gamma_q = \sum_{p=1}^{P} \gamma_{pq}$ is the combined reflectivity of all scatterers within the patch. For the covariance, the combining matrix for the q-th clutter patch is given by $\mathbf{B}_{\mathbf{q}} = [\mathbf{v}(f_{\mathtt{c},q}) \otimes \mathbf{s} \otimes \mathbf{a}(\theta_q, \phi_q), \ldots, \mathbf{v}(f_{\mathtt{c},q}) \otimes \mathbf{s} \otimes \mathbf{a}(\theta_q, \phi_q)]$. Since the deterministic patch response is repeated P times, via standard Kronecker product properties, this is equivalent to

$$\mathbf{B}_{\mathbf{q}} = \mathbf{1}_P^T \otimes \mathbf{v}(f_{\mathtt{c},q}) \otimes \mathbf{s} \otimes \mathbf{a}(\theta_q, \phi_q). \tag{10}$$

More importantly, since $\mathbf{R}_{\boldsymbol{\gamma}}^{q} = \mathbf{B}_q \mathbf{R}_{\gamma}^{pq} \mathbf{B}_q^H$, we have

$$\mathbf{R}_{\gamma}^{q} = \overline{R}_{\gamma}^{q} (\mathbf{v}_q \otimes \mathbf{s} \otimes \mathbf{a}_q)(\mathbf{v}_q \otimes \mathbf{s} \otimes \mathbf{a}_q)^H \tag{11}$$

where $\overline{R}_{\gamma}^{q} = \mathbf{1}_P^T \mathbf{R}_{\gamma}^{pq} \mathbf{1}_P$.

Relationship to the Channel Model construct: In [18], the authors considered a transfer function/matrix approach for simultaneous transmit and receive resource design in MIMO radar, similar to the typical literature on control theory and digital communications. Using the result

$$(\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C})(\mathbf{D} \otimes \mathbf{E} \otimes \mathbf{F}) = \mathbf{AD} \otimes \mathbf{BE} \otimes \mathbf{CF},$$

we can reframe our model as in [18]. For example, we can write

$$\begin{aligned} \mathbf{v}(f_{\mathtt{c},q}) \otimes \mathbf{s} \otimes \mathbf{a}(\theta_q, \phi_q) &= (\mathbf{v}(f_{\mathtt{c},q}) \otimes \mathbf{I}_N \otimes \mathbf{a}(\theta_q, \phi_q))\mathbf{s} \\ &= \boldsymbol{\Gamma}_q \mathbf{s} \end{aligned}$$

where $\boldsymbol{\Gamma}_q$ is implicitly defined. By the same token, the target response is equivalent to $\mathbf{Ts}$ where $\mathbf{T} = \mathbf{v}_t(f_d) \otimes \mathbf{I}_N \otimes \mathbf{a}_t(\theta_t, \phi_t)$. With this form, the overall received vector in the range gate of interest is

$$\tilde{\mathbf{y}} = \rho_t \mathbf{Ts} + \sum_{q=1}^{Q} \gamma_q \boldsymbol{\Gamma}_q \mathbf{s} + \mathbf{y}_{\mathtt{n}} + \mathbf{y}_{\mathtt{i}}.$$

III. Joint Waveform-Filter Design

Our goal is to find a STAP beamformer vector $\mathbf{w} \in \mathbb{C}^{NML}$ and a transmit signal $\mathbf{s} \in \mathbb{C}^{N}$ that minimizes the combined effect of the noise and interference represented by the covariance matrix $\mathbf{R}_{\mathbf{ni}} \in \mathbb{C}^{NML \times NML}$ and the signal-dependent clutter.

At the range gate we interrogate (which we assume contains the target), the return $\tilde{\mathbf{y}}$ is processed by a filter characterized by the weight vector $\mathbf{w}$, forming the output return $\mathbf{w}^H \tilde{\mathbf{y}}$. As mentioned above, we want to design this vector and the signal to minimize the expected undesired power $\mathbf{E}\{|\mathbf{w}^H \mathbf{y}_{\mathtt{u}}|^2\} = \mathbf{w}^H \mathbf{R}_{\mathtt{u}}(\mathbf{s})\mathbf{w}$. Additionally, we would like to constrain this minimization to ensure reasonable radar operation. First, for a given target space-time-doppler bin, we want a particular filter output, say, $\kappa \in \mathbb{C}$. This filter output can be represented by the Capon beamformer equation $\mathbf{w}^H \mathbf{T}\mathbf{s}$, where $\mathbf{T} \in \mathbb{C}^{NML \times N}$ is the target channel response defined above. Second, we place an upper bound on the total transmit signal power, P_o. Mathematically, we can represent this optimization problem as

$$\begin{aligned} \min_{\mathbf{w},\mathbf{s}} \quad & \mathbf{w}^H \mathbf{R}_{\mathtt{u}}(\mathbf{s})\mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}^H \mathbf{T}\mathbf{s} = \kappa \\ & \mathbf{s}^H \mathbf{s} \le P_o. \end{aligned} \tag{12}$$

In [3], this problem was numerically shown to be non-convex, and therefore a variety of alternating minimization algorithms based on the eigenstructure were used to solve it. In [21], we directly proved (13) was non-convex by examining the stationary points and Hessian of the unconstrained problem and provided engineering intuition as to why this was so. We summarize this contribution in the following lemma and provide a sketch of the proof:

Lemma 1: Let $(\mathbf{w}_o, \mathbf{s}_o)$ be a joint stationary point of the objective function in (13). The joint trans-receive problem in (13) is non-convex unless, for every clutter patch matrix $\mathbf{\Gamma}_q$ and the noise-interference covariance $\mathbf{R}_{\mathbf{ni}}$,

$$\mathbf{w}_o^H \mathbf{\Gamma}_q \mathbf{s}_o = 0, \quad \mathbf{R}_{\mathbf{ni}} \mathbf{w}_o = \mathbf{0}_{NML \times 1}$$

for every such stationary pair $(\mathbf{w}_o, \mathbf{s}_o)$.

Sketch of proof: An unconstrained optimization problem is convex iff the objective function is convex. Furthermore, a real-valued, twice-continuously-differentiable function is convex iff, when evaluated at every stationary point in the domain, its Hessian is positive semidefinite.

Taking the first derivatives, the stationary points of the objective satisfy the following equation pair:

$$\mathbf{R}_{\mathbf{ni}} \mathbf{w} = -\sum_{q=1}^{Q} \overline{R}_{\gamma}^{q} \left(\mathbf{s}^H \mathbf{\Gamma}_q^H \mathbf{w}\right) \mathbf{\Gamma}_q \mathbf{s}$$

$$\sum_{q=1}^{Q} \overline{R}_{\gamma}^{q} \left(\mathbf{w}^H \mathbf{\Gamma}_q \mathbf{s}\right) \mathbf{\Gamma}_q^H \mathbf{w} = \mathbf{0}_{NML \times 1}.$$

Due to the structure of the Hessian of the objective in (13), it is positive semidefinite at a given pair $(\mathbf{w}, \mathbf{s})$ only if $\mathbf{w}^H \mathbf{\Gamma}_q \mathbf{s} = 0$ for every clutter patch $\mathbf{\Gamma}_q$. On its own, this condition is rather restrictive and requires every stationary pair to span at least a Q-dimensional subspace of the product space $\mathbb{C}^{NML} \times \mathbb{C}^{N}$. If this condition is applied to the definition of the stationary points, then we require $\mathbf{R}_{\mathbf{ni}} \mathbf{w}_o = \mathbf{0}_{NML \times 1}$ which, if $\mathbf{R}_{\mathbf{ni}}$ is full rank, means that the only stationary point is the trivial case of no beamformer at all. If $\mathbf{R}_{\mathbf{ni}}$ is not full rank, then the beamformer of every stationary point must also span the noise-and-interference nullspace.

Neither of these conditions can hold in practice, and so we must conclude that the problem is non-convex. ■

IV. The Biquadratic Program & Relaxations

In this section, we discuss the relatively unexplored area of biquadratic programming (BQP/BiQP) – that is, joint optimization of two multidimensional variables over a cost function that is quadratic in each variable – and its application to fully adaptive radar, which arises from joint signal-beamformer design schemes. The primary theory and major applications of BQP have been described in [15], [16], [22]–[26].

A. BQP for Fully Adaptive Radar

Consider an expanded version of the design problem (13) in the pair $(\mathbf{w}, \mathbf{s})$:

$$\begin{aligned} \min_{\mathbf{w},\mathbf{s}} \quad & \mathbf{w}^H \mathbf{R}_{\mathbf{ni}} \mathbf{w} + \mathbf{w}^H \left(\sum_{q=1}^{Q} \overline{R}_{\gamma}^{q} \mathbf{\Gamma}_q \mathbf{s}\mathbf{s}^H \mathbf{\Gamma}_q^H \right) \mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}^H \mathbf{T}\mathbf{s} = \kappa \\ & \mathbf{s}^H \mathbf{s} \le P_o. \end{aligned} \tag{13}$$

If we treat the signal and beamformer as a single stacked variable, say $\mathbf{b} = [\mathbf{w}^T \ \ \mathbf{s}^T]^T \in \mathbb{C}^{N(ML+1)}$, we can find an equivalent form of the above optimization problem in the new variable. Define $\mathbf{\Psi}_W = [\mathbf{I}_{NML} \ \mathbf{0}_{NML \times N}]$ and $\mathbf{\Psi}_S = [\mathbf{0}_{N \times NML} \ \mathbf{I}_{NML}]$. Then, we can define the optimization problem as

$$\begin{aligned} \min_{\mathbf{b}} \quad & \mathbf{b}^H \widetilde{\mathbf{R}}_{\mathbf{ni}} \mathbf{b} + \mathbf{b}^H \left(\sum_{q=1}^{Q} \overline{R}_{\gamma}^{q} \overline{\mathbf{\Gamma}}_q \mathbf{b}\mathbf{b}^H \overline{\mathbf{\Gamma}}_q^H \right) \mathbf{b} \\ \text{s.t.} \quad & \mathbf{b}^H \widetilde{\mathbf{T}}\mathbf{b} = \kappa \\ & \mathbf{b}^H \mathbf{\Psi}_S^T \mathbf{\Psi}_S \mathbf{b} \le P_o, \end{aligned} \tag{14}$$

where $\widetilde{\mathbf{R}}_{\mathbf{ni}} = \mathbf{\Psi}_W^T \mathbf{R}_{\mathbf{ni}} \mathbf{\Psi}_W$, $\overline{\mathbf{\Gamma}}_q = \mathbf{\Psi}_W^T \mathbf{\Gamma}_q \mathbf{\Psi}_S$, and $\widetilde{\mathbf{T}} = \mathbf{\Psi}_W^T \mathbf{T} \mathbf{\Psi}_S$ are "expanded" versions of the matrices seen in the previous problem. Both (14) and (15) are BQPs. In fact, (15) is the special case of a quartic optimization problem.

Due to Ling, *et al.* [15], we know a variety of facts about the BQP for real variables: it is nonconvex (though they did not demonstrate any practical implications of the non-convexity as we have), both the BQP and its naive semidefinite relaxation (which we will more explicitly discuss in the next section) are NP-hard problems, and the homogenous general problem does not even admit a polynomial time *approximation* algorithm. While this seems to bode ill for our situation, they also demonstrated that applying a semidefinite relaxation provides a fast, if not perfectly accurate, computational solution method. Additionally, if the structure is sufficiently sparse, existing solvers

can attack the semidefinite relaxation even more efficiently. As we will show later, our structure can and will be significantly sparse, so our method will take advantage of this property.

In either case, since these problems are non-convex, the question becomes: can we find a sufficient convex relaxation to obtain approximate solutions?

B. Semi-Definite Relaxation

Our first goal is using the notation/mechanisms of the literature to find the semidefinite relaxation of the equivalent programs (14) and (15). First, assume we have defined a tensor $\mathcal{C}$ such that the following operator relation holds: $\mathcal{C}\mathbf{s}\mathbf{s}^H = \sum_{q=1}^{Q} \mathbf{R}_{\boldsymbol{\gamma}}^q(\mathbf{s})$. Similarly, assume an expanded form of this tensor exists, say $\widetilde{\mathcal{C}}$, such that $\widetilde{\mathcal{C}}\mathbf{b}\mathbf{b}^H = \sum_{q=1}^{Q} \overline{\boldsymbol{\Gamma}}_q \mathbf{b}\mathbf{b}^H \overline{\boldsymbol{\Gamma}}_q^H$. Based on the structure of this tensor operator, we know that any matricization of it is, in fact, positive semidefinite, which will be important later.

We can recast the complex constraint $\mathbf{w}^H\mathbf{T}\mathbf{s} = \kappa$ as two real constraints that depend on the joint parameter vector $\mathbf{b}$:

$$\begin{aligned}\mathrm{Re}\{\mathbf{w}^H\mathbf{T}\mathbf{s}\} &= \mathbf{b}^H\widetilde{\mathbf{T}}_R\mathbf{b} = \kappa_R\\ \mathrm{Im}\{\mathbf{w}^H\mathbf{T}\mathbf{s}\} &= \mathbf{b}^H\widetilde{\mathbf{T}}_I\mathbf{b} = \kappa_I\end{aligned}$$

where the subscripts R and I indicate the Real and Imaginary parts of the expanded matrix $\widetilde{\mathbf{T}}$, respectively. (As an aside, recall that both matrices are implicitly Hermitian, since their quadratic forms produce real values.) If we use the assumed tensor $\mathcal{C}$ and the properties of the trace, we can recast the quartic optimization problem above into the standard form:

$$\begin{aligned}\min_{\mathbf{b}} \quad & \widetilde{\mathbf{R}}_{\mathbf{ni}} \bullet \mathbf{b}\mathbf{b}^H + \left(\widetilde{\mathcal{C}}\mathbf{b}\mathbf{b}^H\right) \bullet \mathbf{b}\mathbf{b}^H\\ \text{s.t.} \quad & \widetilde{\mathbf{T}}_R \bullet \mathbf{b}\mathbf{b}^H = \kappa_R\\ & \widetilde{\mathbf{T}}_I \bullet \mathbf{b}\mathbf{b}^H = \kappa_I\\ & \boldsymbol{\Psi}_S^T\boldsymbol{\Psi}_S \bullet \mathbf{b}\mathbf{b}^H \le P_o\end{aligned}$$

where $\mathbf{X} \bullet \mathbf{Y} = \mathrm{tr}(\mathbf{X}^H\mathbf{Y})$ indicates the standard inner product on compatible matrices. A common path to a semidefinite relaxation defines a matrix variable $\mathbf{B} = \mathbf{b}\mathbf{b}^H$. It is clear that this Hermitian PSD matrix $\mathbf{B} \in \mathbb{H}_J^+$ is rank one, which imposes an obviously non-convex constraint. If, instead, we permit $\mathbf{B}$ to be of any rank, we have relaxed the problem to the quadratic semidefinite program (QSDP) [27], [28]:

$$\begin{aligned}\min_{\mathbf{B}} \quad & \widetilde{\mathbf{R}}_{\mathbf{ni}} \bullet \mathbf{B} + \left(\widetilde{\mathcal{C}}\mathbf{B}\right) \bullet \mathbf{B}\\ \text{s.t.} \quad & \widetilde{\mathbf{T}}_R \bullet \mathbf{B} = \kappa_R\\ & \widetilde{\mathbf{T}}_I \bullet \mathbf{B} = \kappa_I\\ & \boldsymbol{\Psi}_S^T\boldsymbol{\Psi}_S \bullet \mathbf{B} \le P_o\\ & \mathbf{B} \in \mathbb{H}_J^+.\end{aligned} \tag{15}$$

Notice we did not, at any point, progress through a matrix bilinear form in finding this semidefinite relaxation, as seen in [15], etc. This is because the constraint set in this problem prevents such a representation – namely, the Capon constraint is bilinear in the vector arguments already.

If, as we will show below, the above QSDP is convex, Toh, *et al.* [28] point out that computation of such a problem depends primarily on the number of linear constraints (here, three) and the effective rank of the tensor $\widetilde{\mathcal{C}}$ (upper bounded by the number of clutter patches Q). More specifically, if reformulated into a semidefinite-quadratic-linear program (SQLP), the computational complexity for an interior-point solver is no more than $\mathcal{O}((2(Q+3))^3)$, where the additional factor of two accounts for operating on complex matrices.

If we can obtain a solution to this problem, a natural question is how to apply this solution to the original problem. If the resulting solution is rank one, the problem is directly solved. If the rank is greater than one, we can use the nearest rank-one approximation that still satisfies the constraints. We will see in Section VII below that for this particular scenario, we can obtain solutions that are nearly rank-one in a numerical sense.

V. Solutions of the Relaxed BQP

With the general form of our relaxation now developed, we now turn to examining how best to solve for the relaxed matrix $\mathbf{B}$ and its relation to an optimal beamformer-signal pair $(\mathbf{w}, \mathbf{s})$. While numerical methods will be the primary driver of obtaining such solutions, we can use some analytic techniques to sketch, at the very least, the notional contours of the relaxed solution and its relation to the channel state parameters and signal-independent noise-and-interference covariance. In this section, we begin the task of analyzing the optimal solutions of (16) by first developing a more realizable form of the relaxed optimization discussed in Section IV which we recognize to be convex. Next, we demonstrate that there is no duality gap (and thus have a quite strong convergence guarantee) under a very reasonable, physically-realizable restriction. Finally, we summarize the necessary and sufficient conditions for optimality and provide explicit linkage of these conditions to the STAP problem.

Before we begin, we reformulate the problem to use the variable $\boldsymbol{\beta} = \mathrm{vec}(\mathbf{B})$. So long as we use the appropriate rules of differentiation (see [29]) and remember that there is an isomorphism between Hermitian matrices of size n and the real vector space of size n^2 (which manifests as a matrix operation on $\boldsymbol{\beta}$), this is an appropriate representation for our variable of interest. Additionally, all of our usual inner products, etc. are maintained, so there is no mathematical conflict with this choice.

In order to make use of this vectorization, observe that the tensor $\widetilde{\mathcal{C}}$ can be unfolded into the $J^2 \times J^2$ matrix $\widetilde{\mathbf{C}}_V = \sum_{q=1}^{Q} \overline{R}_q^{\gamma}\, \mathrm{vec}(\overline{\boldsymbol{\Gamma}}_q)\, \mathrm{vec}(\overline{\boldsymbol{\Gamma}}_q)^H$. Consider also that if $\mathbf{C}_V = \sum_{q=1}^{Q} \overline{R}_q^{\gamma}\, \mathrm{vec}(\boldsymbol{\Gamma}_q)\, \mathrm{vec}(\boldsymbol{\Gamma}_q)^H$, then $\widetilde{\mathbf{C}}_V = (\boldsymbol{\Psi}_S \otimes \boldsymbol{\Psi}_W)^T \mathbf{C}_V (\boldsymbol{\Psi}_S \otimes \boldsymbol{\Psi}_W)$. Using these, we can standardize and vectorize the QSDP (16) into the equivalent problem

$$\begin{aligned}\min_{\mathbf{B}\in\mathbb{H}_J^+} \quad & \mathrm{vec}^H(\widetilde{\mathbf{R}}_{\mathbf{ni}})\boldsymbol{\beta} + \boldsymbol{\beta}^H\widetilde{\mathbf{C}}_V\boldsymbol{\beta}\\ \text{s.t.} \quad & \mathrm{vec}^H(\widetilde{\mathbf{T}}_R)\boldsymbol{\beta} - \kappa_R = 0\\ & \mathrm{vec}^H(\widetilde{\mathbf{T}}_I)\boldsymbol{\beta} - \kappa_I = 0\\ & \mathrm{vec}^H(\boldsymbol{\Psi}_S^T\boldsymbol{\Psi}_S)\boldsymbol{\beta} - P_o \le 0.\end{aligned} \tag{16}$$

It is clear that, since the feasible set is the intersection of the PSD cone and two affine constraints and the objective is a non-homogenous convex quadratic function, the vectorized problem (17) (and hence the QSDP (16)) is convex. However, since $\tilde{\mathbf{C}}_V$ is typically *not* full rank (indeed, the rank is bounded above by ML in this case), it is not usually strictly convex, and thus, in general, there will exist a multitude of solutions.

A. Slater's Condition

Recall that *strong duality* is said to hold for a given optimization problem if the primal problem is convex and it satisfies Slater's condition [14, pp. 240, 265] (i.e., the problem is *strictly feasible*). A consequence of this property is that there is no duality gap between the solutions of the primal and dual problems. We will show that our problem satisfies Slater's condition given a relationship between the constraint values.

Given an optimization problem

$$\begin{aligned} \min_{\mathbf{x}\in\Omega} \quad & f_o(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \le 0 \quad i = 1,\dots,m \\ & h_j(\mathbf{x}) = 0 \quad j = 1,\dots,n \end{aligned}$$

where Ω is a convex set, assume $f_o(\mathbf{x})$ is convex and define the set $\mathcal{D} = \Omega \cap \operatorname{dom}(f_o) \cap (\cap_{i=1}^m \operatorname{dom}(g_i))$ to be the total feasible domain of the problem. Slater's condition is satisfied if there exists at least one point in the relative interior (i.e. not on the boundary) of the problem's feasibility set that satisfies all of the equality constraints and strictly satisfies the inequality constraints.

For our problem, the feasibility set is positive semidefinite matrices, whose relative interior is positive-definite matrices. Our equivalent complex constraint functions are

$$\begin{aligned} g(\mathbf{B}) &= \boldsymbol{\psi}_S^T \boldsymbol{\beta} - P_o \\ h_1(\mathbf{B}) &= \boldsymbol{\beta}^H \tilde{\boldsymbol{\tau}} - \kappa \;\; = \tilde{\boldsymbol{\tau}}^T \mathbf{K}_{J,J} \boldsymbol{\beta} - \kappa \\ h_2(\mathbf{B}) &= \tilde{\boldsymbol{\tau}}^H \boldsymbol{\beta} - \kappa^* \end{aligned}$$

where $\tilde{\boldsymbol{\tau}} = \operatorname{vec}(\tilde{\mathbf{T}})$, and $\boldsymbol{\psi}_S = \operatorname{vec}(\boldsymbol{\Psi}_S^T \boldsymbol{\Psi}_S)$. Thus, we need to find a matrix $\mathbf{B} \succ 0$ that satisfies the above equations. We assert that the Slater condition is satisfied for the relaxed BQP given a reasonable condition, which we describe in the following theorem.

Theorem 1: The relaxed joint trans-receive design problem in (17) satisfies Slater's condition if

$$\|\mathbf{T}\|_F^2 > \frac{|\kappa|^2}{P_o}$$

The details of the proof of this theorem can be found in Appendix A. This is indeed a reasonable assumption, because it says that the Capon constraint, i.e. the target-matching goal, cannot exceed the overall reflected power from the target.

In the side-looking STAP case, we can make a more precise relation to the dimensionality of the problem, using the fact that the target matrix has a Kronecker structure – that is, $\mathbf{T} = \mathbf{v}_t \otimes \mathbf{I}_N \otimes \mathbf{a}_t$. This means that $\mathbf{T}^H \mathbf{T} = \|\mathbf{v}_t\|^2 \|\mathbf{a}_t\|^2 \mathbf{I}_N$. If we assume that the spatial response comes from a uniform linear array and that the Doppler response is similarly uniform, then $\|\mathbf{a}_t\|^2 = M$ and $\|\mathbf{v}_t\|^2 = L$. This further implies that $\mathbf{T}^H \mathbf{T} = ML\mathbf{I}_N$ and $\|\mathbf{T}\|_F^2 = NML$. Hence, via the theorem, Slater's condition is satisfied so long as $NML > \frac{|\kappa|^2}{P_o}$.

In either case, since most reasonable STAP problems (or, indeed, any joint trans-receive problem in active sensing) satisfy this condition, we can be assured that almost any numerical method used to solve (16) will converge to a solution that is both primally and dually optimal (or nearly so).

B. Obtaining the KKTs

In the subsequent analysis, let us assume that Theorem 1 holds for our specific problem. It is well-known that, for an optimization problem that satisfies Slater's condition, the Karush-Kuhn-Tucker (KKT) conditions are necessary first-order conditions for the existence and optimality of a set of primal and dual variables. Furthermore, if such a problem is convex, then the KKTs become necessary and sufficient conditions, which implies that solving them directly may provide insight into the solutions of the problem. In this section, we summarize and analyze the KKTs for the problem (17); further details of their derivation can be found in [21] and Appendix B.

Let $\tilde{\mu} \in \mathbb{C}$, $\lambda \in \mathbb{R}$, $\boldsymbol{\Sigma} \in \mathbb{H}_+^J$ be the Lagrange multipliers/dual variables associated with the Capon, power, and semidefiniteness constraints in (17), respectively. Furthermore, let $\boldsymbol{\sigma} = \operatorname{vec}(\boldsymbol{\Sigma})$. Finally, let $\mathcal{H}[\cdot]$ indicate the Hessian of a given function. Then, generically, the KKT conditions for the relaxed biquadratic problem can be written as

1) $\nabla_{\mathbf{B}} \mathcal{L}(\mathbf{B}^o, \tilde{\mu}^o, \lambda^o, \boldsymbol{\Sigma}^o) = \mathbf{0}_{J\times J}$
2) $\lambda^o \ge 0,\ \boldsymbol{\Sigma}^o \succeq 0$
3) $\lambda^o(\operatorname{tr}(\boldsymbol{\Psi}_S^T \mathbf{B}^o \boldsymbol{\Psi}_S) - P_o) = 0,\ \mathbf{B}^o \boldsymbol{\Sigma}^o = \mathbf{0}_{J\times J}$
4) $\operatorname{tr}(\boldsymbol{\Psi}_S^T \mathbf{B}^o \boldsymbol{\Psi}_S) - P_o \le 0,\ \mathbf{B} \succeq 0$
5) $\operatorname{tr}(\mathbf{B}^o \boldsymbol{\Psi}_W^T \mathbf{T} \boldsymbol{\Psi}_S) = \kappa$
6) $\mathcal{H}[\mathcal{L}(\mathbf{B}^o, \tilde{\mu}^o, \lambda^o, \boldsymbol{\Sigma}^o)] \succeq 0$

where the optimal value of a variable is denoted by a superscript o and $\nabla_{\mathbf{B}}$ indicates a gradient on the manifold defined by the domain of the variable $\mathbf{B}$. That is, for a matrix $\mathbf{B}^o$ to be a regular minimizer of the related optimization problem, it is necessary that it satisfies these conditions. If the problem is convex, these are necessary and sufficient conditions for the optimal minimizer and associated Lagrange multipliers. For future notational simplicity, we will drop the superscript o until absolutely necessary (i.e., a final statement of the optimal solution). According to [29], the gradient in the first KKT condition is $\nabla_{\mathbf{B}} = \frac{\partial}{\partial B^*}$ or, in other words, $\mathcal{D}_{\mathbf{B}^*}\mathcal{L} = \operatorname{vec}^T(\frac{\partial \mathcal{L}}{\partial B^*})$. Thus, this vectorized form can take the place of the gradient.

A brief summary of the salient elements of the KKTs follows. First, we note that the Hessian of the Lagrangian is merely $\tilde{\mathbf{C}_V}$, which is PSD by definition, and thus the relevant condition is always satisfied. Next, by rearranging and solving for the gradient condition, we obtain an expression for the semidefiniteness multiplier $\boldsymbol{\Sigma}$ (which itself is PSD):

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{R}_{\mathbf{ni}} & \boldsymbol{\Sigma}_2 \\ \boldsymbol{\Sigma}_2^H & \lambda \mathbf{I}_N \end{bmatrix}, \tag{17}$$

where $\boldsymbol{\Sigma}_2 = \sum_{q=1}^Q \overline{R}_\gamma^q \operatorname{tr}(\boldsymbol{\Gamma}_q^H \mathbf{B}_2)\boldsymbol{\Gamma}_q - \tilde{\mu}^* \mathbf{T}$. Note the dependence of the slackness variable on the optimal solution, which

we have partitioned as $\mathbf{B} = \left[\begin{smallmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{B}_2^H & \mathbf{B}_3 \end{smallmatrix}\right]$, with block sizes equivalent to those of $\mathbf{\Sigma}$. If we vectorize $\mathbf{\Sigma}_2$, we obtain a more obvious connection to the original cost function:

$$\boldsymbol{\sigma}_2 = \mathbf{C}_V \boldsymbol{\beta}_2 - \tilde{\mu}^* \boldsymbol{\tau}. \tag{18}$$

Hence, the magnitude of this dual variable can be regarded as a "remainder" of sorts – that is, the difference between how much the optimal primal solution must align with the clutter (the first term) and the target itself, scaled by the remaining gain from the equality constraint (the second term).

In order for this equation to have a solution for $\boldsymbol{\beta}_2$ (which a non-trivial set of optimal variables will have), we also require

$$\mathbf{P}_{\mathbf{C}_V}^{\perp} \boldsymbol{\sigma}_2 = -\tilde{\mu}^* \mathbf{P}_{\mathbf{C}_V}^{\perp} \boldsymbol{\tau}, \tag{19}$$

that is, the portion of $\mathbf{\Sigma}_2$ with columns orthogonal to each clutter patch is a scaled version of the portion of the target similarly situated. Furthermore, due to the positive semidefinitess of $\mathbf{\Sigma}$, the clutter-target remainder $\mathbf{\Sigma}_2$ must be contained within signal-independent noise-and-interference spectrum, i.e., $\mathcal{R}(\mathbf{\Sigma}_2) \subseteq \mathcal{R}(\mathbf{R_{ni}})$.

Finally, we summarize some results from the application of the matrix complementarity conditions in Appendix B. Namely, we can find two equivalent representations of the multiplier λ

$$\lambda = \frac{\mathrm{tr}(\mathbf{B}_1 \mathbf{R_{ni}})}{P_o} = -\frac{\mathrm{tr}(\mathbf{B}_2^H \mathbf{\Sigma}_2)}{P_o}, \tag{20}$$

as well as a characterization of the objective function in terms of a dual variable,

$$\tilde{\mu}^* \kappa = \boldsymbol{\beta}_2^H \mathbf{C}_V \boldsymbol{\beta}_2 + \mathrm{tr}(\mathbf{B}_1 \mathbf{R_{ni}}). \tag{21}$$

From (20), we have an interpretation of $\frac{1}{\lambda}$ (when λ is positive) as an SINR of sorts; that is, we can regard it as the ratio of the transmit power to the power received from non-signal dependent sources after filtering. More interestingly, we can see from (21) that the optimal rejection of undesired interference and noise is ultimately dependent on the Capon constraint and its dual variable. In fact, this equivalence comes about irrespective of particular constraints on the transmitted signal, which will be useful in future analyses of more realistically-constrained design problems.

VI. Consequences of the KKTs

The optimality conditions shown above and in Appendix B are, at first blush, a complicated set of matrix equations to solve. Furthermore, to our knowledge, there are no generic conditions in the literature that would allow us to immediately identify properties of the relaxed solution or its relationship to high-quality solutions of the particular BQP (14). Indeed, most work on QSDPs (see [30] and citations within) has eschewed such guarantees or only provides them for quadratic operators very much unlike the clutter tensor $\mathcal{C}$. However, we can derive some problem-specific insight by examining and solving for certain KKT conditions that have physical meaning.

In this section, we will first prove a variety of generic properties for any solution $\mathbf{B}^o$ of (16) and then relate these properties to the ability to directly recover high-quality solutions to (14). In particular, we will show that if the signal-independent noise-and-interference covariance matrix $\mathbf{R_{ni}}$ is full rank, then the solution is both power-bounded and more likely to have simple solution recovery. With these properties in hand, we will link our relaxation with historical joint trans-receive design by showing that the power-bounded solution produces a result very similar in spirit to waterfilling – in this case, the "water" is the joint trans-receive resource "filling" (or nulling, as necessary) the combined target-clutter spectrum.

A. General Properties of the Relaxed Solution

Since, in our problem, the power constraint is an inequality, it is worthwhile to ask under what conditions the solution reaches that bound – that is, when does $\mathrm{tr}(\mathbf{B}_3^o) = P_o$? It turns out that this is a pivotal component of characterizing any solution of (16). Not only does this allow us to anticipate the role of clutter whitening in the solution, but it also allows us to predict a bound on the solution matrix's rank.

Before we begin, we note that due to the complementary slackness condition (KKT condition 3 above), the power bound is attained if and only if the dual variable $\lambda > 0$. Therefore, this condition will be shorthand in our proofs for attaining the power bound.

First, we show that the solution does not reach the power bound P_o iff the slackness matrix $\mathbf{\Sigma}_2 = \mathbf{0}_{NML \times N}$ by proving the following lemma.

Lemma 2: $\lambda = 0 \iff \mathbf{\Sigma}_2 = \mathbf{0}_{NML \times N}$

Proof: First, we proceed in the forward direction. If $\lambda = 0$, then the slackness matrix becomes

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{R_{ni}} & \mathbf{\Sigma}_2 \\ \mathbf{\Sigma}_2^H & \mathbf{0}_{N \times N} \end{bmatrix}.$$

To be part of a feasible solution, this must be positive semidefinite, which is only possible if $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_2 (\mathbf{0}_{N \times N})^{\dagger} (\mathbf{0}_{N \times N}) = \mathbf{0}_{NML \times N}$ (see [31, Theorem $I_{a''}$]). Hence, the forward direction is proved.

In the reverse direction, we prove via contradiction. Assume that $\mathbf{\Sigma}_2 = \mathbf{0}_{NML \times N}$ and $\lambda > 0$. The matrix slackness condition 41 dictates that $\lambda \mathbf{B}_2 = -\mathbf{B}_1 \mathbf{\Sigma}_2$. Under our first assumption, this becomes $\lambda \mathbf{B}_2 = \mathbf{0}_{NML \times N}$, which simplifies to $\mathbf{B}_2 = \mathbf{0}_{NML \times N}$ under the second assumption. However, any feasible solution must also satisfy the Capon constraint $\mathrm{tr}(\mathbf{B}_2^H \mathbf{T}) = \kappa \neq 0$. Clearly, $\mathbf{B}_2 = \mathbf{0}_{NML \times N}$ violates this constraint, which leads to our contradiction and completes the proof. ■

Observe that if $\mathbf{\Sigma}_2 = \mathbf{0}_{NML \times N}$, then, via (18), recovery of an optimal $\mathbf{B}_2$ would merely be a whitening of the target with respect to the clutter matrix $\mathbf{C}_V$ (since then $\mathbf{C}_V \boldsymbol{\beta}_2 = \tilde{\mu}^* \boldsymbol{\tau}$). However, as shown above, this is not possible in a power-bounded solution, which means that additional resources in the joint spectrum must have been allocated to some other whitening process.

It turns out that this other whitening process is related to signal-*independent* interference and noise, as we will show below. Indeed, every feasible solution reaches the power bound if and only if the noise-and-interference correlation matrix $\mathbf{R_{ni}}$ is full rank, the proof of which makes use of Lemma 2.

Proposition 1: $\lambda > 0 \iff \mathrm{rank}(\mathbf{R_{ni}}) = NML$

Proof of $\mathrm{rank}(\mathbf{R_{ni}}) = NML \Rightarrow \lambda > 0$*:* We will prove this by contradiction as well. Assume that $\mathbf{R_{ni}}$ is full rank and $\lambda = 0$. Given the second condition, Lemma 2 requires that $\mathbf{\Sigma}_2 = \mathbf{0}_{NML \times N}$. If we apply this to the slackness condition 40, then $\mathbf{B}_1\mathbf{R_{ni}} = \mathbf{0}_{NML \times NML}$. However, since $\mathbf{R_{ni}}$ is full rank, this implies $\mathbf{B}_1 = \mathbf{0}_{NML \times NML}$. This violates our non-triviality, but we will continue with the proof to show we reach a further contradiction. Since the overall solution matrix must be PSD, $\mathbf{B}_2 = \mathbf{0}_{NML \times N}$ as well. As in the proof of Lemma 2, we have reached a contradiction because the Capon constraint is violated, which completes the proof of sufficiency. ■

Proof of $\lambda > 0 \Rightarrow \mathrm{rank}(\mathbf{R_{ni}}) = NML$*:* First, observe that for a non-trivial PSD solution matrix, $\mathcal{R}(\mathbf{B}_2^H) \subseteq \mathcal{R}(\mathbf{B}_3)$ by [32, Theorem 7.7.9(a,b)]. Next, we turn to slackness condition (43), which states $\lambda\mathbf{B}_3 = -\mathbf{B}_2^H\mathbf{\Sigma}_2$. If $\lambda > 0$, then clearly $\mathcal{R}(\mathbf{B}_3) \subseteq \mathcal{R}(\mathbf{B}_2^H)$. By the standard rules of subset inclusion, then, $\mathcal{R}(\mathbf{B}_2^H) = \mathcal{R}(\mathbf{B}_3)$ and $\mathrm{rank}(\mathbf{B}_2^H) = \mathrm{rank}(\mathbf{B}_3)$. This fact will become useful later.

We will now use the results of [33] on another slackness condition (42) and use a substitution of (43) to get to our destination. Using [33, Theorem 2.2] on (42) to solve for $\mathbf{R_{ni}}$, we have the following requirements for $\mathbf{R_{ni}}$ to be at least PSD (which it is):

1) $-\mathbf{B}_3\mathbf{\Sigma}_2^H\mathbf{B}_2 \succeq 0$: If we substitute (43) into this, we obtain $\lambda\mathbf{B}_3^2 \succeq 0$, which is satisfied because $\lambda > 0$ and $\mathbf{B}_3$ is PSD.
2) $\mathcal{R}(\mathbf{B}_3\mathbf{\Sigma}_2^H) \subseteq \mathcal{R}(\mathbf{B}_2^H)$: This is satisfied because $\mathbf{B}_3\mathbf{\Sigma}_2^H = -\frac{1}{\lambda}\mathbf{B}_2^H\mathbf{\Sigma}_2\mathbf{\Sigma}_2^H$ via (43), and the range inclusion follows directly.
3) $\mathrm{rank}(-\mathbf{B}_3\mathbf{\Sigma}_2^H\mathbf{B}_2) = \mathrm{rank}(\mathbf{B}_3\mathbf{\Sigma}_2^H)$: This is not immediately satisfied by the other conditions, but it does imply that $\mathrm{rank}(\mathbf{B}_3) = \mathrm{rank}(\mathbf{B}_3\mathbf{\Sigma}_2^H)$.

The more interesting requirement comes from [34], which adds the following: $\mathbf{R_{ni}}$ is positive definite (and thus full rank) if and only if $\mathrm{rank}(-\mathbf{B}_3\mathbf{\Sigma}_2^H\mathbf{B}_2) = \mathrm{rank}(\mathbf{B}_2^H)$. We know from the third PSD requirement above that $\mathrm{rank}(-\mathbf{B}_3\mathbf{\Sigma}_2^H\mathbf{B}_2) = \mathrm{rank}(\mathbf{B}_3)$, and thus $\mathbf{R_{ni}}$ is full rank if and only if $\mathrm{rank}(\mathbf{B}_3) = \mathrm{rank}(\mathbf{B}_2^H)$. However, as seen from above, this condition is already satisfied if $\lambda > 0$, and so our proof is complete. ■

In cases where there is perfect state information (that is, full knowledge of the noise and interference covariance *a priori*), the noise-and-interference covariance is *always* full-rank, so Proposition 1 allows us to conclude that most theoretical solutions will achieve the power bound. This would also be true if there is sufficient sample support to provide good estimates of the non-clutter/non-target processes. However, computation or insufficent training data might result in a rank-deficient estimate of $\mathbf{R_{ni}}$, which would mean the power bound is not attained. The effects of such a scenario are detailed in [21]. Furthermore, this means that if a rank-one solution is obtained, then the relaxation will have provided us a solution that is directly comparable to power-equality-constrained problems, like those in [10].

Since the proof of Proposition 1 provides us with the evidence to show that power-bounded solutions occur only in non-singular noise-and-interference environments, we can also demonstrate an additional property of any power-bounded solution: namely, the rank of the overall solution matrix. Recall that for $\mathbf{B}$ to be PSD, $\mathcal{R}(\mathbf{B}_2) \subseteq \mathcal{R}(\mathbf{B}_1)$. However, the slackness condition (40) provides that $\mathbf{B}_1\mathbf{R_{ni}} = -\mathbf{B}_2\mathbf{\Sigma}_2^H$. Since $\mathbf{R_{ni}}$ is full rank, (40) becomes $\mathbf{B}_1 = -\mathbf{B}_2\mathbf{\Sigma}_2^H\mathbf{R_{ni}}^{-1}$, which implies $\mathcal{R}(\mathbf{B}_1) \subseteq \mathcal{R}(\mathbf{B}_2)$. Thus, $\mathcal{R}(\mathbf{B}_1) = \mathcal{R}(\mathbf{B}_2)$, and $\mathrm{rank}(\mathbf{B}_1) = \mathrm{rank}(\mathbf{B}_2) = \mathrm{rank}(\mathbf{B}_3) \leq N$, where the last equality is implied by Proposition 1's proof and the inequality is obvious.

In fact, we can strengthen this inequality with the following lemma, which relates the pseudo-SINR λ to the optimal power-bounded rank.

Lemma 3: In a power-bounded solution, $\mathrm{rank}(\mathbf{B}) \leq N - \mathrm{rank}(\lambda\mathbf{I}_N - \mathbf{\Sigma}_2^H\mathbf{R_{ni}}^{-1}\mathbf{\Sigma}_2)$.

Proof: Our above note on Proposition 1 is our starting point. Since the matrix product $\mathbf{B\Sigma}$ is zero, so is its rank. This further implies $\mathrm{rank}(\mathbf{B}) \leq J - \mathrm{rank}(\mathbf{\Sigma})$. In general, since $\mathbf{\Sigma}$ is PSD, $\mathrm{rank}(\mathbf{\Sigma}) = \mathrm{rank}(\mathbf{R_{ni}}) + \mathrm{rank}(\lambda\mathbf{I}_N - \mathbf{\Sigma}_2^H\mathbf{R_{ni}}^{\dagger}\mathbf{\Sigma}_2)$. Therefore, generally, $\mathrm{rank}(\mathbf{B}) \leq (NML - \mathrm{rank}(\mathbf{R_{ni}})) + (N - \mathrm{rank}(\lambda\mathbf{I}_N - \mathbf{\Sigma}_2^H\mathbf{R_{ni}}^{\dagger}\mathbf{\Sigma}_2))$. In a power bounded solution, $\mathrm{rank}(\mathbf{R_{ni}}) = NML$ and the inverse exists, thus $\mathrm{rank}(\mathbf{B}) \leq N - \mathrm{rank}(\lambda\mathbf{I}_N - \mathbf{\Sigma}_2^H\mathbf{R_{ni}}^{-1}\mathbf{\Sigma}_2)$. ■

Hence, the only way to guarantee rank-one solutions (and thus immediate solution to (14)) is for $\mathrm{rank}(\lambda\mathbf{I}_N - \mathbf{\Sigma}_2^H\mathbf{R_{ni}}^{-1}\mathbf{\Sigma}_2) = N - 1$. That said, as we will show below in Section VII, if there is a single dominant eigenvalue, then the effective rank will be one, and a high quality power-bounded solution may be obtained. This proof also implies that if the solution is not power-bounded, then recovery of a high-quality approximation may be more difficult, as the relaxed solution potentially spans a much larger space (up to $NML - \mathrm{rank}(\mathbf{R_{ni}})$ more directions!).

Finally, one might ask if the optimal solution produces non-physical SINR guarantees, since the objective function is the undesired signal variance (and thus the denominator of SINR). For power-bounded solutions, the following lemma confirms that this is impossible; for context, recall from (21) that $\tilde{\mu}^o\kappa$ is the optimal value of (15)/(16).

Lemma 4: If $\lambda > 0$, $\tilde{\mu} \neq 0$.

Proof: We prove by contradiction. Assume $\tilde{\mu} = 0$. When applied to (18), we have $\boldsymbol{\sigma}_2 = \mathbf{C}_V\boldsymbol{\beta}_2$. Premultiplying with $\boldsymbol{\beta}_2^H$, we have $\boldsymbol{\beta}_2^H\boldsymbol{\sigma}_2 = \boldsymbol{\beta}_2^H\mathbf{C}_V\boldsymbol{\beta}_2 \geq 0$ as the quadratic form of a PSD matrix. But we have already established that for $\lambda > 0$, $\boldsymbol{\beta}_2^H\boldsymbol{\sigma}_2 < 0$, which is a contradiction and our proof is complete. ■

Hence, a power-bounded solution will necessarily allow some remaining noise and interference energy, small as it might be. Further insights into possible "complete" nulling for non-power bounded solutions can be found in [21].

B. A Connection with Waterfilling

Using some of these general results from above, we can show that the optimal solution to the KKTs at the power bound satisfies equations that resemble the well-known "waterfilling" concept.

First, we begin with a decomposition of the clutter matrix $\mathbf{C}_V$. Recall that the rank of the clutter "subspace" (i.e., the rank of $\mathbf{R_c}(\mathbf{s})$ and thus $\mathbf{C}_V$) is limited by both the physical extent and nature of non-target scatterers and our overall ability to observe this state of nature. For side-looking airborne arrays, the well-known Brennan rule [5], [35] is a reasonable approximation if certain conditions hold, but as [36] showed, a more robust result is obtained by applying the Landau-Pollak theorem. In any case, let us assume that the "true" rank of the clutter is $Q_{\mathrm{eff}} \leq Q \leq N^2ML$ (the subscript denoting the *effective* number of clutter

patches). Then, we can use the economy eigendecomposition to find two equivalent representations of $\mathbf{C}_V$, namely:

$$\mathbf{C}_V = \breve{\mathbf{U}}_C \mathbf{D}_C \breve{\mathbf{U}}_C^H = \sum_{i=1}^{Q_{\text{eff}}} \nu_i \breve{\mathbf{u}}_i \breve{\mathbf{u}}_i^H.$$

Here, $\breve{\mathbf{U}}_C \in \mathbb{C}^{N^2ML \times Q_{\text{eff}}}$ is the matrix that forms the basis for the Q_{eff}-dimensional vectorized clutter subspace, whose ith column is the eigenvector $\breve{\mathbf{u}}_i \in \mathbb{C}^{N^2ML}$, $i \in \{1, \ldots, Q_{\text{eff}}\}$. $\mathbf{D}_C \in \mathbb{C}^{Q_{\text{eff}} \times Q_{\text{eff}}}$ is a diagonal matrix whose (i,i)th element is the nonzero eigenvalue ν_i. We can also consider the full eigendecomposition $\mathbf{C}_V = \breve{\mathbf{U}}\mathbf{D}\breve{\mathbf{U}}$. Here, the unitary matrix $\breve{\mathbf{U}} = [\breve{\mathbf{U}}_C \ \breve{\mathbf{U}}_N]$, where $\breve{\mathbf{U}}_C$ is as above and $\breve{\mathbf{U}}_N$ collects the $N^2ML - Q_{\text{eff}}$ eigenvectors in the nullspace of $\mathbf{C}_V$, which we can regard as the vectors $\breve{\mathbf{u}}_i, i \in \{Q_{\text{eff}}+1, \ldots, N^2ML\}$. $\mathbf{D}$ is just the direct sum of $\mathbf{D}_C$ and a $N^2ML - Q_{\text{eff}} \times N^2ML - Q_{\text{eff}}$ all-zeros matrix. This decomposition also provides us with an alternative representation of the signal-dependent clutter covariance matrix $\mathbf{R}_{\mathsf{c}}(\mathbf{s})$. Let us assume that there is a set of N^2ML matrices $\mathbf{U}_i \in \mathbb{C}^{NML \times N}$ whose vectorizations are $\text{vec}(\mathbf{U}_i) = \breve{\mathbf{u}}_i$ – that is, they correspond to the eigenvectors of $\mathbf{C}_V$. Then, the signal-dependent clutter covariance can also be given by

$$\mathbf{R}_{\mathsf{c}}(\mathbf{s}) = \sum_{i=1}^{Q_{\text{eff}}} \nu_i \mathbf{U}_i \mathbf{s}\mathbf{s}^H \mathbf{U}_i^H. \tag{22}$$

We note that $\mathbf{U}_i^H \mathbf{U}_i = \frac{1}{N}\mathbf{I}_N$ for all i, not just those in the clutter indices. Thus, $\sqrt{N}\mathbf{U}_i$ is a rank-N partial isometry.

Now, given this formulation, we can show a waterfilling-like effect by combining the matrix slackness condition (42) and one of the Lagrangian conditions. If we recast (18) in matrix form, and use the set of partial isometries, it expands to

$$\mathbf{\Sigma}_2 = \sum_{i=1}^{Q_{\text{eff}}} \nu_i \,\text{tr}(\mathbf{U}_i^H \mathbf{B}_2)\, \mathbf{U}_i - \tilde{\mu}^* \mathbf{T}. \tag{23}$$

Substituting this into (42) gives us

$$\mathbf{R}_{\mathbf{ni}} \mathbf{B}_2 = \tilde{\mu}^* \mathbf{T}\mathbf{B}_3 - \sum_{i=1}^{Q_{\text{eff}}} \nu_i \,\text{tr}(\mathbf{U}_i^H \mathbf{B}_2)\, \mathbf{U}_i \mathbf{B}_3. \tag{24}$$

Let us assume that we are power-bounded and everything that implies from the lemmas above. Thus, we can "directly" find $\mathbf{B}_2$ by applying $\mathbf{R}_{\mathbf{ni}}^{-1}$ above to obtain:

$$\mathbf{B}_2 = \tilde{\mu}^* \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{T}\mathbf{B}_3 - \sum_{i=1}^{Q_{\text{eff}}} \nu_i \,\text{tr}(\mathbf{U}_i^H \mathbf{B}_2)\, \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_i \mathbf{B}_3 \tag{25}$$

Premultiplying both sides with $\mathbf{U}_j^H$ and taking the trace gives us

$$\begin{aligned}\text{tr}(\mathbf{U}_j^H \mathbf{B}_2) &= \tilde{\mu}^* \,\text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{T}\mathbf{B}_3) \\ &\quad - \sum_{i=1}^{Q_{\text{eff}}} \nu_i \,\text{tr}(\mathbf{U}_i^H \mathbf{B}_2)\ \text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_i \mathbf{B}_3).\end{aligned} \tag{26}$$

We can extract the jth term from the sum and collect it on the left hand side to obtain

$$\begin{aligned}&\text{tr}(\mathbf{U}_j^H \mathbf{B}_2)(1 + \nu_j\ \text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_j \mathbf{B}_3)) \\ &\qquad = \tilde{\mu}^* \,\text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{T}\mathbf{B}_3) \\ &\qquad\quad - \sum_{\substack{i=1 \\ i \neq j}}^{Q_{\text{eff}}} \nu_i \,\text{tr}(\mathbf{U}_i^H \mathbf{B}_2)\ \text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_i \mathbf{B}_3).\end{aligned} \tag{27}$$

Now, the second term on the left-hand side of the above equation could be divided out as long as there was a guarantee it was always positive. First, if $\nu_j = 0$, this term is 1, which is positive. Otherwise, $\nu_j > 0$ since they are the non-zero eigenvalues of a positive semidefinite matrix. Next, we need to examine the trace statement. $\mathbf{B}_3$ is positive semidefinite by construction, and $\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_j$ is positive definite because each $\mathbf{U}_j$ is a full rank scaled partial isometry and $\mathbf{R}_{\mathbf{ni}}$ is positive definite (see [32, Theorem 7.7.2]). Thus, the trace of this matrix product is always positive and real. With this in hand, we can now say

$$\begin{aligned}\text{tr}(\mathbf{U}_j^H \mathbf{B}_2) &= \tilde{\mu}^* \frac{\text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{T}\mathbf{B}_3)}{1 + \nu_j\ \text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_j \mathbf{B}_3)} \\ &\quad - \sum_{\substack{i=1 \\ i \neq j}}^{Q_{\text{eff}}} \frac{\nu_i\ \text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_i \mathbf{B}_3)}{1 + \nu_j\ \text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_j \mathbf{B}_3)} \,\text{tr}(\mathbf{U}_i^H \mathbf{B}_2).\end{aligned} \tag{28}$$

Since $\mathbf{B}_3$ is a relaxed version of $\mathbf{s}\mathbf{s}^H$, we can regard the expression $\nu_j\ \text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_j \mathbf{B}_3)$ to be a relaxed form of $\nu_j\ \mathbf{s}^H \mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_j \mathbf{s}$, which is (effectively) a clutter-to-noise-and-interference ratio for the jth basis matrix. Similarly, $\text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{T}\mathbf{B}_3)$ is a joint target-and-clutter to noise-and-interference ratio, and the cross term $\nu_i\ \text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_i \mathbf{B}_3)$ captures the ratio of patch-to-patch interaction and the noise-and-interference level. Thus, the first term on the right hand side is effectively a normalized measure of the clutter-matched target spectrum given a signal basis $\mathbf{B}_3$, and the right hand side is a normalized measure of the intraclutter spectrum given that same basis.

Traditional waterfilling dictates that power is preferentially injected to bands where the overall signal-to-noise ratio is high, proceeding to "worse"-off bands until the available power is exhausted. This is somewhat *inverted* in (28) because the left-hand side is an unscaled version of the subspace alignment between the solution $\mathbf{B}_2$ and the jth canonical clutter transfer matrix $\mathbf{U}_j$. Naively, we would like to *minimize* this alignment for $j \in \{1, \ldots, Q_{\text{eff}}\}$, since aligning with the clutter would nominally degrade our matching of the target. However, the coupled nature of (28) requires more nuance than that. First, observe that for the non-clutter directions, i.e., $j \in \{Q_{\text{eff}}+1, ..., N^2ML\}$, $\nu_j = 0$ and (28) becomes

$$\begin{aligned}\text{tr}(\mathbf{U}_j^H \mathbf{B}_2) &= \tilde{\mu}^* \,\text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{T}\mathbf{B}_3) \\ &\quad - \sum_{i=1}^{Q_{\text{eff}}} \nu_i\ \text{tr}(\mathbf{U}_j^H \mathbf{R}_{\mathbf{ni}}^{-1} \mathbf{U}_i \mathbf{B}_3) \,\text{tr}(\mathbf{U}_i^H \mathbf{B}_2).\end{aligned} \tag{29}$$

This means that the solution's alignment in the non-clutter directions follows that of the whitened target's, less the combined crossover between non-clutter and clutter in the whitened spectrum. If the target is strong in these directions, the solution will align towards them. Otherwise, the solution must match the target in-spectrum as near as possible in directions both where the whitened clutter power is the lowest and the alignment with other directions is minimized. This corresponds to the findings in [19], which showed a similar behavior in two-step mutual information waveform design over consecutive transmit epochs. In this case, the "two steps" can be regarded as the trans-receive pair instead of sequential temporal designs, and the "separation" is between target and clutter instead of different targets.

But what of the value $\tilde{\mu}$? This clearly relates to the total available resources – in this case, κ and P_o – and provides us with the "water" in waterfilling. We can find a form of $\tilde{\mu}$ by reexamining (25) as follows. First, we premultiply by the target matrix $\mathbf{T}^H$ and take the trace of both sides. Recognizing that a feasible solution satisfies $\mathrm{tr}(\mathbf{T}^H\mathbf{B}_2) = \kappa^*$, we now have

$$\begin{aligned}\kappa^* &= \tilde{\mu}^* \,\mathrm{tr}(\mathbf{T}^H\mathbf{R}_{\mathbf{ni}}^{-1}\mathbf{T}\mathbf{B}_3) \\ &\quad - \sum_{i=1}^{Q_{\mathrm{eff}}} \nu_i \,\mathrm{tr}(\mathbf{U}_i^H\mathbf{B}_2)\ \mathrm{tr}(\mathbf{T}^H\mathbf{R}_{\mathbf{ni}}^{-1}\mathbf{U}_i\mathbf{B}_3). \qquad (30)\end{aligned}$$

This is where the traditional waterfilling appears, since we are saying that the gain across the target (which we know in this case to be κ^*) is the upper bound on the available resources (given by the first term) minus the overall impact of the clutter weighted by its alignment (the second term). According to the above argument, $\mathrm{tr}(\mathbf{T}^H\mathbf{R}_{\mathbf{ni}}^{-1}\mathbf{T}\mathbf{B}_3)$ is never zero. Therefore, after rearrangement, a final form for $\tilde{\mu}^*$ is:

$$\tilde{\mu}^* = \frac{\kappa^* + \sum_{i=1}^{Q_{\mathrm{eff}}} \nu_i \,\mathrm{tr}(\mathbf{U}_i^H\mathbf{B}_2)\ \mathrm{tr}(\mathbf{T}^H\mathbf{R}_{\mathbf{ni}}^{-1}\mathbf{U}_i\mathbf{B}_3)}{\mathrm{tr}(\mathbf{T}^H\mathbf{R}_{\mathbf{ni}}^{-1}\mathbf{T}\mathbf{B}_3)} \qquad (31)$$

Thus, the optimal solution to the relaxed problem (particularly when power-bounded) describes a generalized whiten-and-match trans-receive filter process that exhibits waterfilling behavior, shaping the transmit process to match the target's response in clutter as much as possible while still minimizing the clutter response.

VII. Simulations and Results

In this section, we demonstrate the utility of our proposed relaxation scheme through comparative numerical simulation. While we have preliminarily reduced solving the QSDP (16) to a matrix completion problem, it is possible to solve the problem numerically with commercial solvers. Most of the analysis presented here is enabled by the modeling toolbox CVX [37], [38] and the solvers SDPT3 [39], [40] and SeDuMi[41].

Unless otherwise stated, the scenarios presented to the solver were as follows:

As noted in [21], our available computational resources and current simulation environment constrain our overall problem size. Therefore, in this case, we assume $N = 5$, $M = 5$, and $L = 16$. However, we do not believe that this is an inherent limitation which precludes the applicability of our technique to larger problem sizes.

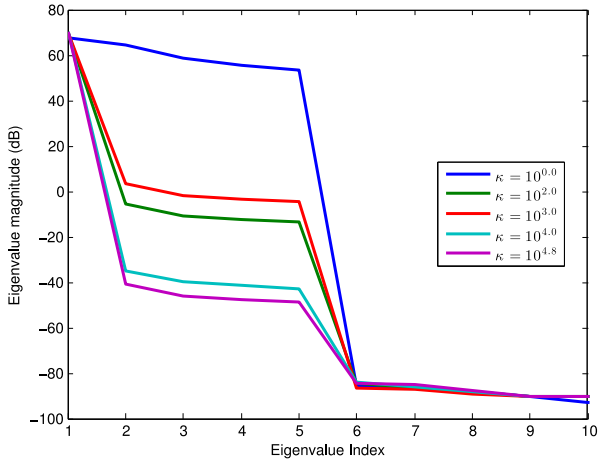
The radar operates on a carrier frequency of $f_o = 1$ GHz and transmits pulses at a pulse repetition frequency $f_p = 20$ KHz. The receive array has an interelement spacing of $d = \lambda_o/2$. The noise covariance matrix was a scaled correlation matrix with correlation function $\exp(-0.05|n|)$ for $n \in \{1, \ldots, NML\}$. Interferers were placed at the azimuth-elevation pairs (0.3941, 0.3) radians and $(-0.4951, 0.3)$ radians, with correlation matrices given by the Toeplitz matrix associated with the correlation function $\exp(0.2|l|)$ for $l \in \{1, \ldots, NL\}$. The clutter was modeled with $Q = 25$ patches of $P = 5$ scatterers each, equally spaced over the azimuth interval $(-\pi/2, \pi/2)$ at an elevation angle of $\frac{\pi}{4}$ radians. Adjacent scatterers within each patch are correlated with coefficient -0.2. The target was placed at the angle coordinates $(\theta_t, \phi_t) = (0.3, \frac{\pi}{3})$ and travels at a relative normalized Doppler frequency of -0.255. In all cases, we have made the Capon constraint real.

To obtain $\mathbf{s}$, $\mathbf{w}$ from our method, we generated the best "rank-1" approximation through the principal singular vector of $\mathbf{B}$ weighted by the associated singular value, then rescaled as necessary to meet the constraints. In practice, violations of the equality constraint are minimal and do not affect the overall performance of our method.

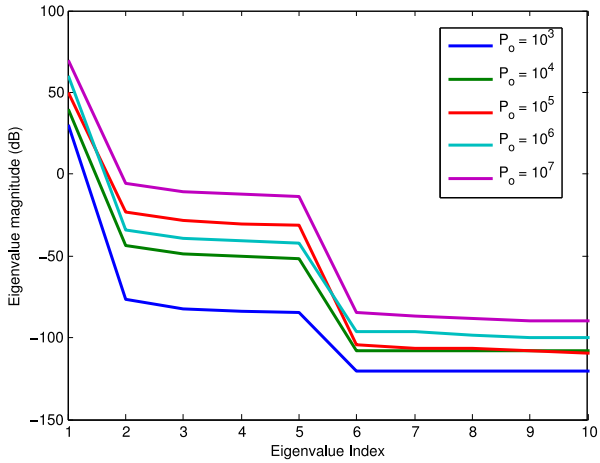
Some of the following results involve comparisons with competing algorithms, which we detail here. First, we consider the alternating minimization (AM) scheme of Setlur & Rangaswamy [3], which Tang and Tang [12] later reported as their "Algorithm 1" despite operating on a different cost function. This technique initializes the problem in (13) with a fixed signal $\mathbf{s}_{init}$ and solves for $\mathbf{w}$ (the W-step), then uses that optimal $\mathbf{w}$ in (13) to solve for a new $\mathbf{s}$ (the S-step), and so on until convergence. We also compare against a modified version of the relaxed alternating minimization scheme of Aubry, *et al*. [10] (AA2), which attempts to directly maximize the SINR through a similar procedure to AM, but uses a linear SDP in the S-step. Our modification removes the similarity constraint, which necessitates a small change in the rank-1 signal decomposition – the similarity constraint matrix is replaced by the target gain matrix. It is here we note two primary advantages of our method: it requires no problem-dependent initialization, other than what the available solving method dictates, and it is not an explicitly iterative scheme.

A. Characterization of the Relaxed Solution

We begin by briefly examining the resulting solution matrix $\mathbf{B}^o$ of the relaxed design problem in (16) and related formulations for the common scenario. Despite QSDPs lacking the solution rank guarantees of linear SDPs, we have found that, depending on the parameter scaling, the given solver produces solutions of rank no greater than N, as predicted by Lemma 3. This rank property of $\mathbf{B}^o$ is demonstrated in Fig. 1, which displays the solution's first $2N$ eigenvalues on a logarithmic scale over variation in κ and P_o for the common scenario above. Fig. 1(a) demonstrates that generally, as κ decreases for a fixed power level (here, $P_o = 10^7$), the effective numerical rank of $\mathbf{B}_o$ approaches N. This indicates a transition region must exist where the desired gain falls below some threshold dictated by the noise, interference, and clutter characteristics. In contrast, as



(a) Capon constraint κ, $P_o = 10^7$



(b) Power constraint P_o, $\kappa = 100$.

Fig. 1. First $2N$ eigenvalues of relaxed solution for varying (a) Capon constraint κ, (b) power constraint P_o.

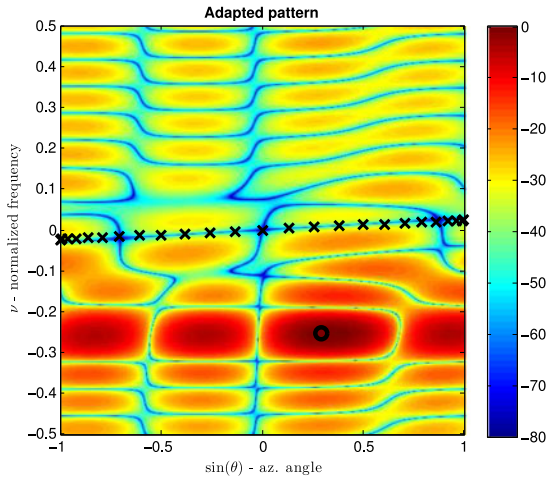


Fig. 2. Adapted Pattern (dB relative to peak), RBQP Solver. Target at ◯, Clutter phase centers at ×. No Interference.

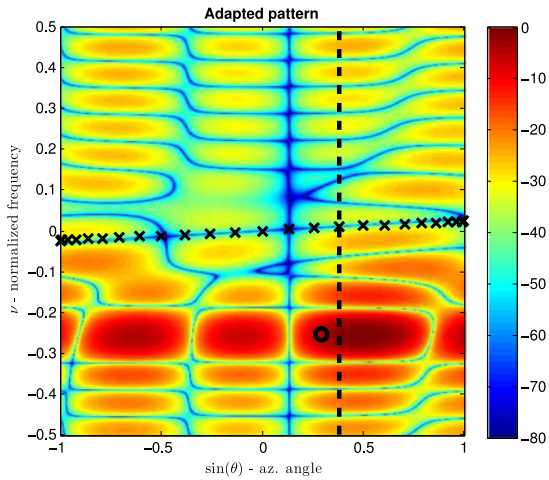


Fig. 3. Adapted Pattern (dB relative to peak), RBQP Solver. Target at ◯, Clutter phase centers at ×, Interferer (dashed line) at $(\theta, \phi) = (0.3941, 0.3)$ radians.

shown in Fig. 1(b), varying the power P_o for a fixed κ (here, $\kappa = 100$) only affects the overall size of the eigenvalues, not the rank. As demonstrated in [21], this also has implications for the utility of constructed rank one approximations. In summary, we have demonstrated that across most parameter choices, the relaxed solution is effectively rank one, which validates our choice of approximate retrieval scheme.

B. Interference Effects

In this subsection, we consider the impact of interference on the relaxed solution only. Our goal here is to determine if the relaxation produces some unknown or unexpected bias or inability to null or mitigate the impact of interference due to the relaxation itself.

To examine this effect, we consider the traditional adapted pattern for STAP, which plots the following function

$$\mathcal{P}(f_d, \theta;\ \phi) = |\mathbf{w}_o^H (\mathbf{v}(f_d) \otimes \mathbf{s}_o \otimes \mathbf{a}(\theta, \phi)|^2. \qquad (32)$$

That is, the adapted pattern for a given beamformer-signal pair $\mathbf{w}_o, \mathbf{s}_o$ is a function of the Doppler frequency f_d and the azimuth θ at a given elevation ϕ. We consider the same simulation parameters as above, with and without interference, and provide the overall adapted pattern at an elevation cut matched to the target.

With no interference, the naive implementation performs relatively well, as illustrated in Fig. 2. The target is well localized, with the peak of the pattern at its location, and a deep null is steered along the clutter ridge (represented by an X at each patch's angle-doppler phase center).

Now suppose we inject a broadband interferer "close" to the target – as a reminder, the target is at $(\theta_t, \phi_t) = (0.3, \frac{\pi}{3})$ radians and the interferer is at $(\theta_I, \phi_I) = (0.3941, 0.3)$ radians. Fig. 3 shows the adapted pattern under these conditions. Note that a clear bias is introduced by the interferer, since in this case we lack the spatial receive resources necessary to null it exactly. Observe, however, that an aliased null is present in the spectrum, indicating that it is merely a sampling issue and not a failure of the scheme to identify the interference. Furthermore, we have found that if the elevation cut of the adapted pattern is taken

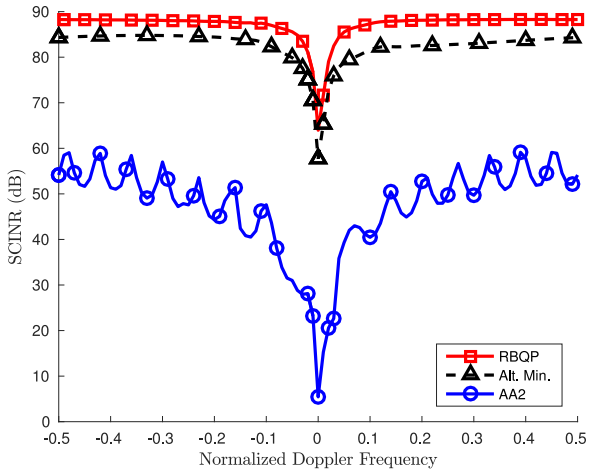


Fig. 4. SINR as a function of target Doppler frequency.

at the elevation of the interferer, a clear null at the appropriate azimuth appears.

We note that this interference impact (in all cases) is mitigated by increasing M in our simulations. This is not surprising, since more antenna elements over the same aperture increases the degrees of freedom available to null interference and localize the target.

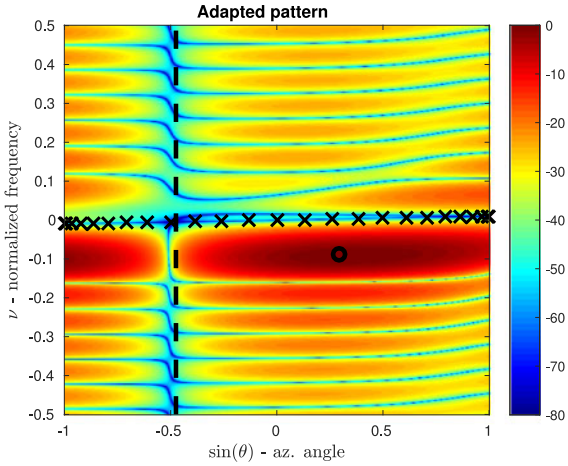
C. Comparison with Alternating Minimization Schemes

In this subsection, we compare the performance of our approximate relaxed BQP (RBQP)/QSDP solution to the existing methodologies in the literature described above.

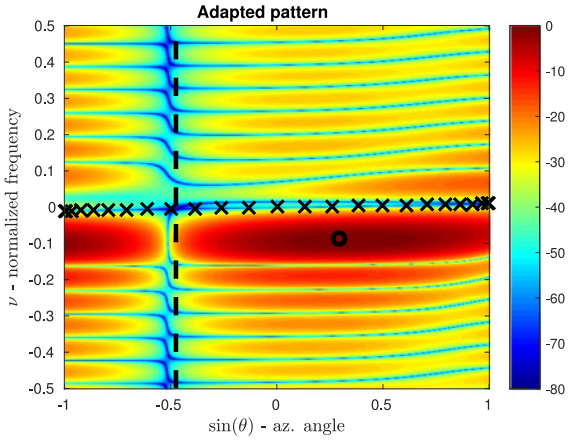
One means of comparison between these algorithms is the robustness of the output SINR for each scheme to variations in the target Doppler frequency f_d. For ease of comparison, we presented the two iterative solvers with parameters we found to assist their convergence: each was initialized with the same chirp waveform having a time-bandwidth product of 50, but the waveform used to initialize AA2 was scaled to have $P_o = 1$, and the resulting SINR was then rescaled to match the actual power constraint $P_o = 10^7$. This is because the authors in [10] claim the algorithm is invariant to scaling. To further facilitate equal comparison, we set $\kappa = \sqrt{P_o}$, which is what AA2 natively produces after rescaling. The convergence threshold for AA2 is set to $\epsilon = 10^{-3}$, and AM is terminated after 20 iterations (which provided a comparable convergence factor). Fig. 4 shows this metric for each of the considered algorithms – our scheme (labeled RBQP), AM, and AA2 – when applied to the common scenario mentioned previously.

First, it is clear that the solutions produced by RBQP and AM are extremely robust to changes in target Doppler at scale, while AA2 has significant variation. Furthermore, in this scenario, our method generally outperforms both competing algorithms in terms of overall SINR and minimum detectable velocity. The relative gap between RBQP and AM at larger Doppler shifts makes sense, given that our method is a relaxation of that in [3] and that AM is terminated rather quickly.

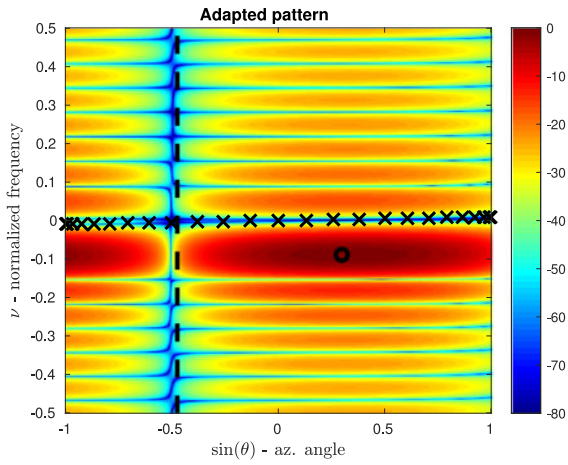
We note here that additional simulations have shown AA2 can outperform both the RBQP procedure and AM, but this requires a number of highly controlled and generally unrealistic assumptions. Namely, the convergence threshold must be fairly high (say, $\epsilon = 1$, if not higher) and $P_o = 1$. Even under these conditions, AA2 often merely produces a permuted version of the initializing waveform, which does not significantly change the initial objective value.



(a) Relaxed BQP (RBQP)



(b) Alt. Min. (AM)



(c) Alg. 2 of [10] (AA2)

Fig. 5. Adapted Patterns (dB rel. to peak) for (a) RBQP, (b) AM, (c) AA2. Target at ◯, Clutter phase centers at ×, Interferer (dashed line) at $(\theta, \phi) = (-0.4951, 0.3)$ radians.

TABLE I
RUNTIME COMPARISON OF ALGORITHMS

Alg.	Med. runtime (sec)	St. Dev. (sec)
RBQP	93.09	4.394
AM	30.95	N/A
AA2	286.53	793.17

The performance benefits provided by RBQP can also be observed in the adapted patterns mentioned in the subsection above. Consider a case where only the second interferer, located at the azimuth-elevation pair $(-0.4951, 0.3)$ radians, is active. Furthermore, assume that the clutter and target are both at the same elevation angle as the interferer. The target remains at an azimuth of 0.3 radians but, due to the elevation change, the target relative Doppler is now -0.087. We note that the only major difference from the previous scenarios is the change in elevation angle for the clutter patches and target.

Fig. 5 shows the adapted pattern at the target elevation cut for each algorithm, with the 0 dB reference set at the notional peak for each algorithm – by construction, this is $|\kappa|^2 = P_o$. However, as we showed in Fig. 4, this does not mean that the algorithms have no difference in detection properties, since the SINR obtained by RBQP is superior to the other two algorithms. Therefore, the adapted patterns shown should not be viewed as directly comparable in terms of SINR, but in terms of relative features and cancellation near the interference, clutter, and target. In all patterns, we can see that the target is well-resolved, deep nulls are placed along the clutter ridge, and a null is placed near the azimuth of the interferer. We can see, though, that both our method (Fig. 5(a)) and AM (Fig. 5(b)) further shape the spectrum by shifting the phase centers of the target response's Doppler sidelobes away from that of the known target (demarcated by the $\bigcirc$) and reducing other sidelobes nearer to the ridge. The relative similarity between these two adapted patterns is generally unsurprising, given the similar SINR performance shown above.

In contrast, AA2 (Fig. 5(c)) retains a more typical grating lobe structure, with no additional nulling or shaping beyond the clutter ridge and interference null. This, in turn, corresponds to the lower SINR seen in Fig. 4. There are a variety of possible explanations for this behavior, the specifics of which are the subject of ongoing research that will be reported in the future.

Not only is a performance improvement obtained, but the RBQP algorithm has a reduced computational load compared with AA2 in [10], as shown in Table I, which lists the median runtime and standard deviation for each algorithm across the Doppler frequency sweep above.

While our algorithm is slower than AM (for a fixed number of iterations), AA2 is inconsistent and slower than either RBQP or AM. In fact, AA2's execution time varied significantly with the target Doppler's relative alignment to the clutter. Again, the restrictive cases mentioned above can improve the runtime of AA2, but at the cost of the aforementioned caveats.

VIII. CONCLUSIONS

In this paper, we reconsidered the problem of [3] under the relaxed biquadratic program framework. We provided a brief non-convexity proof and demonstrated the linkage to QSDPs when relaxed. We then showed that the KKTs require power-bounded solutions when the noise-and-interference matrix is full rank, and that such a solution admits a waterfilling interpretation. Simulations demonstrated that our method produces nearly rank-one solutions in many scenarios and is capable of producing excellent adaptive solutions with a reasonable addtional computational cost. Future work will attempt to generalize these findings and demonstrate its utility for other radar system models that admit a channel response representation. Additionally, we anticipate that exploiting additional realistic waveform constraints will produce even more robust and practical solutions. Finally, we anticipate that exploiting other QSDP solvers will result in better computational costs and more robust solutions for fully-adaptive radar.

APPENDIX A
PROOF OF THEOREM 1

In this appendix, we provide an abbreviated proof of Theorem 1. In order to prove that the Slater Condition is satisfied, we first find the generic matrix that satisfies the equality constraints, then ensure the resulting matrix is both positive definite and strictly satisfies the power inequality. As a combined matrix-vector equation, the complex equality constraints are given by

$$\begin{bmatrix} \tilde{\boldsymbol{\tau}}^T \mathbf{K}_{J,J} \\ \tilde{\boldsymbol{\tau}}^H \end{bmatrix} \boldsymbol{\beta} = \mathbf{E}\boldsymbol{\beta} = \begin{bmatrix} \kappa \\ \kappa^* \end{bmatrix} \tag{33}$$

where $\mathbf{E}$ is implicitly defined.

We first demonstrate that a generic matrix solution exists to this equation and provide its form. Since $\mathbf{E}$ is a fat matrix, its pseudoinverse is $\mathbf{E}^\dagger = \mathbf{E}^H(\mathbf{E}\mathbf{E}^H)^\dagger$. It can be shown that the Gramian matrix above is

$$\begin{bmatrix} \tilde{\boldsymbol{\tau}}^T \mathbf{K}_{J,J} \\ \tilde{\boldsymbol{\tau}}^H \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\tau}}^T \mathbf{K}_{J,J} \\ \tilde{\boldsymbol{\tau}}^H \end{bmatrix}^H = \|\mathbf{T}\|_F^2 \mathbf{I}_2.$$

So long as we have a non-zero target matrix, this is always invertible. Furthermore, this implies that $\mathbf{P}_\mathbf{E} = \mathbf{E}\mathbf{E}^\dagger = \mathbf{I}_2$ and hence a solution to (33) always exists. Therefore, a general solution to the equality constraints is the matrix

$$\mathbf{B} = \mathbf{Z} + \frac{1}{\|\mathbf{T}\|_F^2} \begin{bmatrix} \mathbf{0}_{NML \times NML} & (\kappa^* - \mathrm{tr}(\mathbf{T}^H \mathbf{Z}_2))\mathbf{T} \\ (\kappa - \mathrm{tr}(\mathbf{Z}_2^H \mathbf{T}))\mathbf{T}^H & \mathbf{0}_{N \times N} \end{bmatrix}.$$

where $\mathbf{Z}$ is a Hermitian matrix of identical dimension and partitioning to $\mathbf{B}$. This additional matrix is necessary because the min-norm solution obtained by setting $\mathbf{Z} = \mathbf{0}_{J \times J}$,

$$\mathbf{B}_{\min} = \frac{1}{\|\mathbf{T}\|_F^2} \begin{bmatrix} \mathbf{0}_{NML \times NML} & \kappa^* \mathbf{T} \\ \kappa \mathbf{T}^H & \mathbf{0}_{N \times N} \end{bmatrix},$$

is an indefinite matrix, and hence not a Slater point. Hence, it is now sufficient to prove there exists a $\mathbf{Z}$ such that $\mathbf{B} \succ 0$ and $\mathrm{tr}(\mathbf{Z}_3) < P_o$. We make the judicious choice that $\mathbf{Z}_2 = \mathbf{0}_{NML \times N}$. Under this assumption, our solution matrix is

$$\mathbf{B} = \begin{bmatrix} \mathbf{Z}_1 & \frac{\kappa^*}{\|\mathbf{T}\|_F^2} \mathbf{T} \\ \frac{\kappa}{\|\mathbf{T}\|_F^2} \mathbf{T}^H & \mathbf{Z}_3 \end{bmatrix}.$$

The positive definiteness requirement can then be expressed as the matrix inequalities

$$\mathbf{Z}_1 \succ 0, \quad \mathbf{Z}_3 - \frac{|\kappa|^2}{\|\mathbf{T}\|_F^4}\mathbf{T}^H\mathbf{Z}_1^{-1}\mathbf{T} \succ 0$$

Here, we rely on another judicious choice, setting $\mathbf{Z}_1 = \mathbf{I}_{NML}$, which is clearly positive definite.[1] Now, it is merely sufficient to prove there exists $\mathbf{Z}_3$ such that $\mathbf{Z}_3 \succ \frac{|\kappa|^2}{\|\mathbf{T}\|_F^4}\mathbf{T}^H\mathbf{T}$ *and* $\mathrm{tr}(\mathbf{Z}_3) < P_o$. Let us assume that such a matrix exists. Following [32, Corollary 7.7.4(d)], since $\mathbf{Z}_3 \succ \frac{|\kappa|^2}{\|\mathbf{T}\|_F^2}\mathbf{T}^H\mathbf{T}$, then $\mathrm{tr}(\mathbf{Z}_3) > \frac{|\kappa|^2}{\|\mathbf{T}\|_F^4}\,\mathrm{tr}(\mathbf{T}^H\mathbf{T}) = \frac{|\kappa|^2}{\|\mathbf{T}\|_F^2}$. But we already know that $\mathrm{tr}(\mathbf{Z}_3) < P_o$. Hence, we have the chained inequality

$$P_o > \mathrm{tr}(\mathbf{Z}_3) > \frac{|\kappa|^2}{\|\mathbf{T}\|_F^2}.$$

It therefore follows that there will exist such a matrix $\mathbf{Z}_3$ if and only if $P_o > \frac{|\kappa|^2}{\|\mathbf{T}\|_F^2}$. Hence, we have found the sufficent condition of Theorem 1 directly and the proof is complete. We note that a sufficently good choice is $\mathbf{Z}_3 = \frac{\alpha}{N}\mathbf{I}_N$, with α in the aforementioned interval. ■

APPENDIX B
FURTHER DETAILS OF THE PRIMAL KKTS

In this appendix, we summmarize the derivation of the KKT conditions mentioned in Section V-B.

1) KKT Condition 1: Let $\mathbf{\Sigma}$ be the slackness variable associated with the PSD condition on $\mathbf{B}$, and let $\boldsymbol{\sigma}$ be its vectorization. Furthermore, let $\boldsymbol{\rho} = \mathrm{vec}(\widetilde{\mathbf{R}}_{\mathbf{ni}})$. When vectorized, the Lagrangian of (17) under complex constraints is

$$\mathcal{L}(\mathbf{B}, \mathbf{\Sigma}, \mu, \lambda) = \boldsymbol{\beta}^H(\widetilde{\mathbf{C}}_V\boldsymbol{\beta} + \boldsymbol{\rho} - \tilde{\mu}^*\tilde{\boldsymbol{\tau}} + \lambda\boldsymbol{\psi}_S - \boldsymbol{\sigma}) + \tilde{\mu}^*\kappa - \lambda P_o. \tag{34}$$

After taking the appropriate derivative, KKT Condition 1 is

$$(\widetilde{\mathbf{C}}_V + \mathbf{K}_{J,J}\widetilde{\mathbf{C}}_V^*\mathbf{K}_{J,J})\boldsymbol{\beta} = \boldsymbol{\sigma} - (\boldsymbol{\rho} + \tilde{\mu}^*\tilde{\boldsymbol{\tau}} + \tilde{\mu}\tilde{\boldsymbol{\tau}}_H + \lambda\boldsymbol{\psi}_S). \tag{35}$$

where $\tilde{\boldsymbol{\tau}}_H = \mathrm{vec}(\widetilde{\mathbf{T}}^H)$. Given four disjoint projection matrices $\mathbf{P}_{WW}, \mathbf{P}_{SW}, \mathbf{P}_{WS}, \mathbf{P}_{SS}$ of appropriate dimension, we can decompose the vectorizations of $\mathbf{B}$ and $\mathbf{\Sigma}$ into sums of vectorizations of each partition. For example,

$$\boldsymbol{\beta} = \mathbf{P}_{WW}^T\boldsymbol{\beta}_1 + \mathbf{P}_{SW}^T\boldsymbol{\beta}_2 + \mathbf{P}_{WS}^T\boldsymbol{\beta}_{2,H} + \mathbf{P}_{SS}^T\boldsymbol{\beta}_3$$

where the subscript indicates which submatrix is vectorized, and $\boldsymbol{\beta}_{2,H} = \mathrm{vec}(\mathbf{B}_2^H)$. It can be shown [21], then, that (35) implies (17) so long as

$$\mathbf{C}_V\boldsymbol{\beta}_2 = \boldsymbol{\sigma}_2 + \tilde{\mu}^*\boldsymbol{\tau} \tag{36}$$

whose solution exists if and only if $\mathbf{P}_{\mathbf{C}_V}^{\perp}\boldsymbol{\sigma}_2 = -\tilde{\mu}^*\mathbf{P}_{\mathbf{C}_V}^{\perp}\boldsymbol{\tau}$ where $\mathbf{P}_{\mathbf{C}_V}^{\perp} = \mathbf{I}_{N^2ML} - \mathbf{C}_V\mathbf{C}_V^{\dagger}$ is the orthogonal projector onto the nullspace of $\mathbf{C}_V$. Additionally, the matrices $\mathbf{B}_1$ and $\mathbf{B}_3$ are "free" parameters under (35), but they are constrained by later conditions.

[1]We can make this choice because $NML > N$. If, for whatever reason, the number of transmit resources were greater than the number of receive resources, then we could start with $\mathbf{Z}_3$ and continue from there.

2) KKT Conditions 2-4: The Inequality Constraints: With the gradient condition exhausted, we turn to the inequality constraints (power bound, positive-semidefiniteness of the solution) and their related conditions. For convenience, we shall attack these somewhat independently in separate subsections, though they will interact.

a) The power constraint: Using the partition mentioned above, the conditions related to the power constraint are:

$$\lambda \geq 0 \tag{37}$$

$$\mathrm{Tr}(\mathbf{B}_3) - P_o \leq 0 \tag{38}$$

$$\lambda(\mathrm{Tr}(\mathbf{B}_3) - P_o) = 0. \tag{39}$$

A nice result of the slackness condition is $\lambda\,\mathrm{Tr}(\mathbf{B}_3) = \lambda P_o$, which we will use later. The other "result" is that $\lambda = 0$ when the solution does not reach the power bound, and $\lambda > 0$ when it does. This fact will inform interpretations of the solution found in Section VI.

b) Positive semidefiniteness of $\mathbf{B}$*:* The conditions for semidefiniteness of the relaxed beamformer-signal basis are slightly more complex, but reveal a significant amount of structure to the solution. First, the direct form of these conditions are

$$\mathbf{\Sigma} \succeq 0, \quad \mathbf{B} \succeq 0, \quad \mathbf{\Sigma}\mathbf{B} = \mathbf{0}_{J\times J}.$$

Of course, in this form, they are not especially useful. However, given (17), we know the structure of $\mathbf{\Sigma}$ somewhat. The most useful characterization of semidefiniteness here is that $\mathbf{\Sigma}$ is PSD iff a contraction $\mathbf{X} \in \mathbb{C}^{NML\times N}$ exists such that $\mathbf{\Sigma}_2 = \sqrt{\lambda}\mathbf{R}_{\mathbf{ni}}^{1/2}\mathbf{X}$ (see [32, Theorem 7.7.9(a,b)]).

The complementary slackness condition for the matrix case reduces to the following 4 equalities:

$$\mathbf{B}_1\mathbf{R}_{\mathbf{ni}} = -\mathbf{B}_2\mathbf{\Sigma}_2^H \tag{40}$$

$$\mathbf{B}_1\mathbf{\Sigma}_2 = -\lambda\mathbf{B}_2 \tag{41}$$

$$\mathbf{B}_2^H\mathbf{R}_{\mathbf{ni}} = -\mathbf{B}_3\mathbf{\Sigma}_2^H \tag{42}$$

$$\mathbf{B}_2^H\mathbf{\Sigma}_2 = -\lambda\mathbf{B}_3. \tag{43}$$

These conditions have significant consequences in Section VI, but they also result in equivalent forms for λ. First, taking the trace of (43) and substituting in (39) on the right hand side, we have

$$\lambda = -\frac{\mathrm{tr}(\mathbf{B}_2^H\mathbf{\Sigma}_2)}{P_o}.$$

Since λ is both real and nonnegative, this means that $\mathrm{tr}(\mathbf{B}_2^H\mathbf{\Sigma}_2)$ is real and nonpositive. We can apply this to the trace of (40) to obtain another form, $\lambda = \frac{\mathrm{tr}(\mathbf{B}_1\mathbf{R}_{\mathbf{ni}})}{P_o}$.

We can also apply this logic to (36) to reveal an interesting consequence about the cost function. If premultiplied by $\boldsymbol{\beta}_2$ and rearranged,

$$\boldsymbol{\beta}_2^H\boldsymbol{\sigma}_2 = \boldsymbol{\beta}_2^H\mathbf{C}_V\boldsymbol{\beta}_2 - \tilde{\mu}^*\boldsymbol{\beta}_2^H\boldsymbol{\tau}$$

Applying the equality constraint, this is equivalent to

$$\boldsymbol{\beta}_2^H\boldsymbol{\sigma}_2 = \boldsymbol{\beta}_2^H\mathbf{C}_V\boldsymbol{\beta}_2 - \tilde{\mu}^*\kappa.$$

(Incidentally, this implies $\tilde{\mu}^*\kappa$ is real, and hence the optimal phase of $\tilde{\mu}$ is that of κ.) From the above, however, we can see that $-\boldsymbol{\beta}_2^H \boldsymbol{\sigma}_2 = \lambda P_o = \mathrm{tr}(\mathbf{B}_1 \mathbf{R}_{\mathbf{ni}})$, and so

$$\tilde{\mu}^*\kappa = \boldsymbol{\beta}_2^H \mathbf{C}_V \boldsymbol{\beta}_2 + \mathrm{tr}(\mathbf{B}_1 \mathbf{R}_{\mathbf{ni}}). \tag{44}$$

The right hand side of (44) is immediately recognizable as our objective function, which implies a potential future analysis on the dual problem that is equivalent to our findings here.

3) KKT Condition 5: Equality Constraints: This is the final major KKT condition left to examine, because KKT Condition 6 is trivially satisfied by $\mathbf{C}_V$ being positive semidefinite. Here, our primary concern is the equality constraint $\boldsymbol{\beta}_2^H \boldsymbol{\tau} = \kappa$. According to the first KKT condition, we know that

$$\boldsymbol{\beta}_2 = \mathbf{C}_V^{\dagger}(\boldsymbol{\sigma}_2 + \tilde{\mu}^* \boldsymbol{\tau}) + \mathbf{P}_{\mathbf{C}_V}^{\perp} \mathbf{z}_2. \tag{45}$$

for some vector $\mathbf{z}_2 \in \mathbb{C}^{N^2 ML}$. Substituting this into the equality constraint gives us

$$\boldsymbol{\sigma}_2^H \mathbf{C}_V^{\dagger} \boldsymbol{\tau} + \tilde{\mu} \boldsymbol{\tau}^H \mathbf{C}_V^{\dagger} \boldsymbol{\tau} + \mathbf{z}_2^H \mathbf{P}_{\mathbf{C}_V}^{\perp} \boldsymbol{\tau} = \kappa. \tag{46}$$

More importantly, we recognize that the columns of $\mathbf{B}_2$ and $\mathbf{T}$ must align for there to be a non-trivial feasible solution.

REFERENCES

[1] N. L. Owsley, "A recent trend in adaptive spatial processing for sensor arrays: Constrained adaptation," in *Proc. Signal Process.*, 1972, pp. 591–604.

[2] P. Woodward, *Probability and Information Theory, With Applications to Radar*. London, U.K.: Pergamon, 1953.

[3] P. Setlur and M. Rangaswamy, "Waveform design for radar STAP in signal dependent interference," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 19–34, Jan. 2016.

[4] R. Klemm, *Principles of Space-Time Adaptive Processing*. London, U.K.: Institution of Electrical Engineers, 2002.

[5] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Lab., Massachusetts Inst. of Technol., Lexington, KY, USA, Tech. Rep. ESC-TR-94-109, 1994.

[6] D. Madurasinghe and A. P. Shaw, "Mainlobe jammer nulling via TSI finders: A space fast-time adaptive processor," *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 1, 2006, Art. no. 048789.

[7] Y. Seliktar, D. B. Williams, and E. J. Holder, "A space/fast-time adaptive monopulse technique," *EURASIP J. Appl. Signal Process.*, vol. 2006, 2006, Art. no. 14510.

[8] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Jan. 1969.

[9] A. Aubry, A. DeMaio, M. Piezzo, A. Farina, and M. Wicks, "Cognitive design of the receive filter and transmitted phase code in reverberating environment," *IET Radar Sonar Navig.*, vol. 6, no. 9, pp. 822–833, Dec. 2012.

[10] A. Aubry, A. DeMaio, 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, no. 1, pp. 93–117, Jan. 2013.

[11] C. Y. Chen and P. P. Vaidyanathan, "MIMO radar waveform optimization with prior information of the extended target and clutter," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3533–3544, Sep. 2009.

[12] B. Tang and J. Tang, "Joint design of transmit waveforms and receive filters for MIMO radar space-time adaptive processing," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4707–4722, Sep. 2016.

[13] 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.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[15] C. Ling, J. Nie, L. Qi, and Y. Ye, "Biquadratic optimization over unit spheres and semidefinite programming relaxations," *SIAM J. Optim.*, vol. 20, pp. 1286–1310, 2009.

[16] C. Ling, X. Zhang, and L. Qi, "Semidefinite relaxation approximation for multivariate bi-quadratic optimization with quadratic constraints," *Numer. Linear Algebra Appl.*, vol. 19, no. 1, pp. 113–131, Apr. 2011. [Online]. Available: http://dx.doi.org/10.1002/nla.781

[17] J. R. Magnus and H. Neudecker, "The Commutation Matrix: Some Properties and Applications," *Ann. Statist.*, vol. 7, no. 2, pp. 381–394, Mar. 1979.

[18] J. R. Guerci, J. S. Bergin, R. J. Guerci, M. Khanin, and M. Rangaswamy, "A new MIMO clutter model for cognitive radar," in *Proc. IEEE Radar Conf.*, May 2016, pp. 1–6.

[19] P. Setlur, N. Devroye, and M. Rangaswamy, "Radar waveform design with the two step mutual information," in *Proc. IEEE Radar Conf.*, May 2014, pp. 1317–1322.

[20] J. Guerci and E. Baranoski, "Knowledge-aided adaptive radar at DARPA: An overview," *IEEE Signal Process. Mag.*, vol. 23, no. 1, pp. 41–50, Jan. 2006.

[21] S. M. O'Rourke, P. Setlur, M. Rangaswamy, and A. L. Swindlehurst, "Relaxed bi-quadratic optimization for joint filter-signal design in signal-dependent space-time adaptive processing (STAP)," , Air Force Res. Lab., Sensors Directorate, Interim AFRL-RY-WP-TR-2016–0197, Dec. 2016. [Online]. Available: https://arxiv.org/abs/1703.08115

[22] I. M. Bomze, C. Ling, L. Qi, and X. Zhang, " Standard bi-quadratic optimization problems and unconstrained polynomial reformulations," *J. Global Optim.*, vol. 52, no. 4, pp. 663–687, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10898–011-9710-5

[23] Y. Yang and Q. Yang, " On solving biquadratic optimization via semidefinite relaxation," *Comput. Optim. Appl.*, vol. 53, no. 3, pp. 845–867, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10589–012-9462-2

[24] Y. Yang, Q. Yang, and L. Qi, "Approximation bounds for trilinear and biquadratic optimization problems over nonconvex constraints," *J. Optim. Theory Appl.*, vol. 163, no. 3, pp. 841–858, Dec. 2014.

[25] X. Zhang, C. Ling, and L. Qi, "Semidefinite relaxation bounds for bi-quadratic optimization problems with quadratic constraints," *J. Global Optim.*, vol. 49, pp. 293–311, 2011.

[26] S.-L. Hu and Z.-H. Huang, "Alternating direction method for bi-quadratic programming," *J. Global Optim.*, vol. 51, no. 3, pp. 429–446, Nov. 2011.

[27] M. Kojima, S. Shindoh, and S. Hara, "Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices," *SIAM J. Optim.*, vol. 7, no. 1, pp. 86–125, 1997. [Online]. Available: http://dx.doi.org/10.1137/S1052623494269035

[28] K. Toh, R. Tütüncü, and M. Todd, "Inexact primal-dual path-following algorithms for a special class of convex quadratic SDP and related problems," *Pacific J. Optim.*, vol. 3, pp. 135–164, 2007.

[29] A. Hjørungnes, *Complex-Valued Matrix Derivatives*, 1st ed. Cambridge, UK: Cambridge Univ. Press, 2011.

[30] X. Li, "A two-phase augmented lagrangian method for convex composite quadratic programming," Ph.D. dissertation, Dept. Math., Nat. Univ. Singapore, Singapore, 2015.

[31] E. Kreindler and A. Jameson, "Conditions for nonnegativeness of partitioned matrices," *IEEE Trans. Automat. Control*, vol. 17, no. 1, pp. 147–148, Feb. 1972.

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2013.

[33] C. G. Khatri and S. K. Mitra, " Hermitian and nonnegative definite solutions of linear matrix equations," *SIAM J. Appl. Math.*, vol. 31, no. 4, pp. 579–585, 1976.

[34] J. K. Baksalary, "Nonnegative definite and positive definite solutions to the matrix equation AXA* = B," *Linear Multilinear Algebra*, vol. 16, no. 1–4, pp. 133–139, 1984. [Online]. Available: http://dx.doi.org/10.1080/03081088408817616

[35] L. E. Brennan and F. M. Staudaher, "Subclutter visibility demonstration," Adaptive Sensors Incorporated, Newcastle upon Tyne, U.K., Tech. Rep. RL-TR-92–21, Mar. 1992.

[36] N. A. Goodman and J. M. Stiles, "On clutter rank observed by arbitrary arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 178–186, Jan. 2007.

[37] M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming, version 2.1," Jun. 2015. [Online]. Available: http://cvxr.com/cvx

[38] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Proc. Recent Adv. Learn. Control*, 2008, pp. 95–110. [Online]. Available: http://stanford.edu/ boyd/graph_dcp.html

[39] K. Toh, M. Todd, and R. Tütüncü, "SDPT3—A MATLAB software package for semidefinite programming," *Optim. Methods Softw.*, vol. 11, pp. 545–581, 1999.

[40] R. Tütüncü, K. Toh, and M. Todd, "Solving semidefinite-linear-quadratic programs using SDPT3," *Math. Program., B*, vol. 95, pp. 189–217, 2003.
[41] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11–12, pp. 625–653, 1999. [Online]. Available: http://fewcal.kub.nl/sturm

Sean M. O'Rourke (SM'05–M'15) received the B.S. degree *cum laude* in electrical engineering in 2008 and the M.S. degree in electrical and computer engineering in 2011, both from the University of California, Irvine (UCI), Irvine, CA, USA, where he is currently working toward the Ph.D. degree.

He is currently a Research Electronics Engineer for the Sensors Directorate in the Air Force Research Laboratory (AFRL), Wright-Patterson AFB, OH, USA, where he has been since 2015. Previously, he was a Graduate Researcher at UCI and an ATR Center Intern at Wright State University, Dayton, OH, USA. His research interests include statistical signal processing, particularly for fully adaptive radar, multisensor data fusion and control, tracking systems, optimization theory, and their application to sports analytics.

He received the 2013 US Department of Defense's Science, Mathematics, and Research for Transformation Scholarship. He has also received a student paper award at the 2012 IEEE Sensor Array and Multichannel Signal Processing Workshop. He is a member of Eta Kappa Nu and Tau Beta Pi.

Pawan Setlur (M'12) has more than 12 years experience in radar signal processing, spanning several areas of radar, signal processing, and information theory research, such as, waveform design, radar coding, statistical estimation & detection, radar imaging, and adaptive radar. He is the author of more than 50 peer reviewed IEEE publications. He is a Research Electronics Engineer at Sensors Directorate, AFRL, WPAFB. He is part of the AFOSR Star team, working on the fully adaptive radar paradigm. In 2016, he won the IEEE Dayton Aerospace and Electronic Systems Society award. He has several individual technical accomplishments in areas of radar, signal processing, and information theory. For the past several years, he has been a reviewer for several the IEEE TRANSACTIONS and several IEEE conferences. His graduate and post doctorate research has been funded by DARPA, ONR, ARL, NSF, AFOSR, AFRL, and DRDC Canada.

Muralidhar Rangaswamy (S'89–M'93–SM'98–F'06) received the B.E. degree in electronics engineering from Bangalore University, Bangalore, India, in 1985 and the M.S. and Ph.D. degrees in electrical engineering from Syracuse University, Syracuse, NY, USA, in 1992.

He is currently the Senior Advisor for radar research at the RF Exploitation Branch within the Sensors Directorate, Air Force Research Laboratory (AFRL), Wright-Patterson AFB, OH, USA. Prior to this, he has held industrial and academic appointments. He has coauthored more than 200 refereed journal and conference record papers in the areas of his research interests. Additionally, he is a contributor to eight books and is a coinventor of three U.S. patents. His research interests include radar signal processing, spectrum estimation, modeling non-Gaussian interference phenomena, and statistical communication theory. He received the IEEE Warren White Radar Award in 2013, the 2013 Affiliate Societies Council Dayton (ASC-D) Outstanding Scientist and Engineer Award, the 2007 IEEE Region 1 Award, the 2006 IEEE Boston Section Distinguished Member Award, and the 2005 IEEE-AESS Fred Nathanson memorial outstanding young radar engineer award. He received the 2012 and 2005 Charles Ryan Basic Research award from the Sensors Directorate, AFRL, in addition to more than 40 scientific achievement awards. He is actively involved with the IEEE Aerospace and Electronic Systems Society in a myriad of leadership roles. He worked as the Technical Program Chairman for the 2014 IEEE Radar Conference.

A. Lee Swindlehurst (SM'83–M'84–SM'89–F'04) received the B.S., summa cum laude, and M.S. degrees in electrical engineering from Brigham Young University, Provo, UT, USA, in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1991.

From 1986 to 1990, he was employed at ESL, Inc., Sunnyvale, CA, USA, where he was involved in the design of algorithms and architectures for several radar and sonar signal processing systems. He was the faculty in the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT, USA, from 1990 to 2007, where he was a Full Professor and worked as Department Chair from 2003 to 2006. During 1996–1997, he held a joint appointment as a visiting scholar at both Uppsala University, Uppsala, Sweden, and at the Royal Institute of Technology, Stockholm, Sweden. From 2006 to 2007, he was on leave working as Vice President of Research for ArrayComm LLC, San Jose, CA, USA. He is currently a Professor in the Electrical Engineering and Computer Science Department, University of California Irvine (UCI), Irvine, CA, USA, a former Associate Dean for Research and Graduate Studies in Henry Samueli School of Engineering at UCI, and a Hans Fischer Senior Fellow in the Institute for Advanced Studies, Technical University of Munich, Munich, Germany. His research interests include sensor array signal processing for radar and wireless communications, detection and estimation theory, and system identification, and he has more than 275 publications in these areas. He was Past Secretary of the IEEE Signal Processing Society, past an Editor-in-Chief for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and past member of the Editorial Boards for the *EURASIP Journal on Wireless Communications and Networking*, the IEEE SIGNAL PROCESSING MAGAZINE, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He received the several paper awards: the 2000 IEEE W. R. G. Baker Prize Paper Award, the 2006 and 2010 IEEE Signal Processing Society's Best Paper Awards, the 2006 IEEE Communications Society Stephen O. Rice Prize in the Field of Communication Theory, and is coauthor of a paper that received the IEEE Signal Processing Society Young Author Best Paper Award in 2001.