Widely-linear precoders and decoders for MIMO channels

Donatella Darsena

Università Parthenope, Napoli, Italy,
Email: darsena@uniparthenope.it

Giacinto Gelli and Francesco Verde

Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione
Università Federico II, Napoli, Italy, Email: [gelli, f.verde]@unina.it

***Abstract*—This paper deals with the problem of jointly designing optimum widely-linear (WL) precoders and decoders for multiple-input multiple-output (MIMO) channels, by adopting as a specific design target the minimization of the symbol mean square error and assuming that channel state information is available at both the transmitter and receiver sides. In the proposed scheme, vectors of symbols are precoded through a WL mapping, which jointly operates on the symbol block and its complex-conjugate version; additionally, a WL receiver processes both the received signal and its complex-conjugate counterpart. By using a complex-valued representation of the relevant vectors involved in the design of the WL-MIMO transceiver, we show how the optimization framework originally developed for linear precoding/decoding MIMO structures can be extended to the case where the transmitted symbols and/or the disturbance exhibit improper or noncircularity features.**

I. Introduction

Multiple-input multiple-output (MIMO) communication systems with channel state information (CSI) at both sides of the link can achieve information rate and/or diversity gains [1]. To derive such benefits, a powerful approach consists of using *linear* MIMO transceivers, encompassing linear precoding at the transmitter and linear decoding at the receiver [2]–[6]. The design of linear MIMO transceivers has been typically carried out by using complex-valued vector models, under the assumption that the transmitted symbols and the disturbance are proper [7] or circular [8]. However, when the random vectors involved in system design are improper or noncircular [9], [10], complex linear processing is suboptimal and it is necessary to consider either *augmented real linear* filtering, where each n-dimensional complex vector is represented by an augmented $2n$-dimensional real block by stacking its real and imaginary parts, or *widely-linear (WL)* complex processing [11]–[14], where each n-dimensional complex vector is augmented to a $2n$-dimensional complex block by stacking the vector itself and its complex-conjugate version.

The augmented real linear filtering approach has been adopted in [15] to design MIMO baseband transceivers, under the minimum-mean-square-error (MMSE) criterion, when the transmitted symbols and/or the disturbance exhibit improper or noncircularity features; the obtained strictures exhibit considerable gains over the corresponding complex linear precoding/decoding structures. However, in general, the use of a complex-valued notation to design and analyse a baseband communication system is always preferred, since it models the system in a simpler and more compact way, and, moreover, it allows to provide deeper insight into the obtained solutions. Recently, WL complex techniques have been employed in [16] to jointly design beamforming and combining vectors when improper symbols are transmitted through a MIMO channel corrupted by improper disturbance.

In this paper, by resorting to the MMSE criterion, we use the complex notation framework to optimize WL-MIMO baseband transceivers, which are composed of a WL precoder at the transmitter and a WL decoder at the receiver. The results developed herein generalize and subsume as a particular case the corresponding complex linear MMSE precoding and decoding solutions. The proposed complex-valued framework could be a useful tool to gain further insights into the performance comparison between WL precoding/decoding MIMO techniques and their corresponding linear counterparts.

II. Centrohermitian 2-by-2 blockwise matrices

By extending some results derived in [18], we report basic properties of a 2-by-2 block matrix having the form[1]

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^* & \mathbf{A}_{11}^* \end{bmatrix} \in \mathbb{C}^{2n \times 2m} \qquad (1)$$

where $\mathbf{A}_{11}, \mathbf{A}_{12} \in \mathbb{C}^{n \times m}$. First, we give some definitions.

The matrix $\mathbf{P}_n \in \mathbb{R}^{2n \times 2n}$ defined as

$$\mathbf{P}_n \triangleq \begin{bmatrix} \mathbf{O}_{n \times n} & \mathbf{I}_n \\ \mathbf{I}_n & \mathbf{O}_{n \times n} \end{bmatrix} \qquad (2)$$

[1]The fields of complex and real numbers are denoted with $\mathbb{C}$ and $\mathbb{R}$, respectively; matrices [vectors] are denoted with upper [lower] case boldface letters (e.g., $\mathbf{A}$ or $\mathbf{a}$); the field of $m \times n$ complex [real] matrices is denoted as $\mathbb{C}^{m \times n}$ [$\mathbb{R}^{m \times n}$], with $\mathbb{C}^m$ [$\mathbb{R}^m$] used as a shorthand for $\mathbb{C}^{m \times 1}$ [$\mathbb{R}^{m \times 1}$]; the superscripts $*$, T, H, and -1 denote the conjugate, the transpose, the conjugate transpose, and the inverse of a matrix, respectively; $\Re\{a\}$ and $\Im\{a\}$ are the real and imaginary part of $a \in \mathbb{C}$, respectively; for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ with (i, ℓ) entry $a_{i\ell}$, the (i, ℓ) element of $\mathbf{A}_+$ is $(a_{i\ell})_+ \triangleq \max(a_{i\ell}, 0)$; $\text{rank}(\mathbf{A})$, $\det(\mathbf{A})$, and $\text{tr}(\mathbf{A})$ denote the rank, the determinant, and the trace of $\mathbf{A} \in \mathbb{C}^{n \times n}$; for any $\mathbf{a} \in \mathbb{C}^n$, $\|\mathbf{a}\|$ denotes the Euclidean norm; $\mathbf{O}_{m \times n} \in \mathbb{R}^{m \times n}$ and $\mathbf{I}_m \in \mathbb{R}^{m \times m}$ denote the null and the identity matrices, respectively; $\nabla_{\mathbf{A}^*}(\cdot)$ represents the complex gradient operator with respect to $\mathbf{A}^* \in \mathbb{C}^{n \times m}$; $\mathbf{J}_n \triangleq [\mathbf{I}_n, \mathbf{O}_{n \times n}] \in \mathbb{R}^{n \times 2n}$ is a selection matrix; $j \triangleq \sqrt{-1}$ is the imaginary unit and the operator $\mathbb{E}[\,\cdot\,]$ denotes ensemble averaging.

Also with Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), Research Unit of Napoli. This work was partially supported by the Italian National Project "Harbour Traffic Optimization System" (HABITAT).

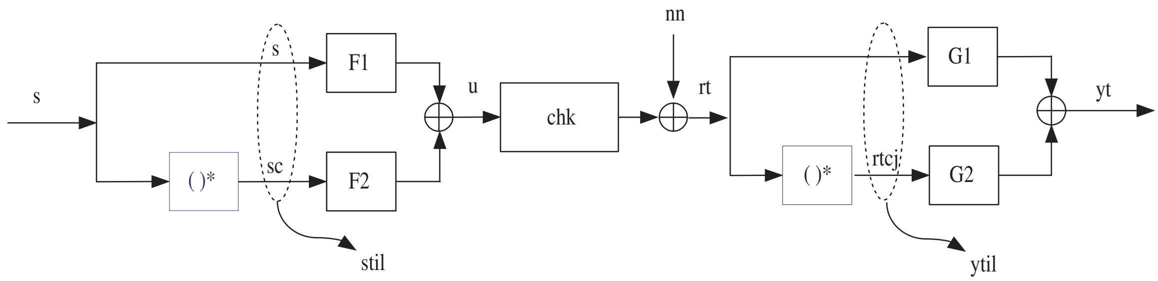

Figure 1. Block diagram of a MIMO communication system with WL precoding and decoding.

is a permutation matrix (i.e., $\mathbf{P}_n^2 = \mathbf{I}_{2n}$ and $\mathbf{P}_n^{\mathrm{T}} = \mathbf{P}_n$). A vector $\tilde{\mathbf{a}} \in \mathbb{C}^{2n}$ is *conjugate symmetric* if $\tilde{\mathbf{a}} = \mathbf{P}_n\, \tilde{\mathbf{a}}^*$. It is clear that, if $\tilde{\mathbf{a}}$ is conjugate symmetric, then $\tilde{\mathbf{a}} = [\mathbf{a}^{\mathrm{T}}, \mathbf{a}^{\mathrm{H}}]^{\mathrm{T}}$, with $\mathbf{a} \in \mathbb{C}^n$. The block matrix $\tilde{\mathbf{A}}$ in (1) exhibits the property $\tilde{\mathbf{A}} = \mathbf{P}_n\, \tilde{\mathbf{A}}^*\, \mathbf{P}_m$ or, equivalently, $\tilde{\mathbf{A}}\, \mathbf{P}_m = \mathbf{P}_n\, \tilde{\mathbf{A}}^*$ and, thus, we say that $\tilde{\mathbf{A}}$ is a *centrohermitian* 2*-by-*2 *blockwise* matrix. The subset of all $2n \times 2m$ centrohermitian 2-by-2 blockwise matrices is denoted by $\mathbb{CH}^{2n\times 2m} \subset \mathbb{C}^{2n\times 2m}$.

Let $\tilde{\mathbf{A}}$ in (1) be square (i.e., $n = m$). The nonsingularity of $\mathbf{A}_{11}$ is a necessary (but not sufficient) condition for the positive definiteness of $\tilde{\mathbf{A}}$. Indeed, if $\mathbf{A}_{11}$ is nonsingular, the Schur complement of $\mathbf{A}_{11}$ in $\tilde{\mathbf{A}}$ is denoted and defined by $\mathbf{A}_{11}/\tilde{\mathbf{A}} \triangleq \mathbf{A}_{11}^* - \mathbf{A}_{12}^*\, \mathbf{A}_{11}^{-1}\, \mathbf{A}_{12}$, and it results [17] that $\det(\tilde{\mathbf{A}}) = \det(\mathbf{A}_{11})\, \det(\mathbf{A}_{11}/\tilde{\mathbf{A}})$. Thus, the matrix $\tilde{\mathbf{A}}$ is nonsingular if and only if (iff) both $\mathbf{A}_{11}$ and $\mathbf{A}_{11}/\tilde{\mathbf{A}}$ are nonsingular, i.e., $\det(\mathbf{A}_{11}) > 0$ and $\det(\mathbf{A}_{11}/\tilde{\mathbf{A}}) > 0$. In this case, by observing that $\mathbf{P}_n^{-1} = \mathbf{P}_n$, one has

$$\tilde{\mathbf{A}}^{-1} = (\mathbf{P}_n\, \tilde{\mathbf{A}}^*\, \mathbf{P}_n)^{-1} = \mathbf{P}_n\, (\tilde{\mathbf{A}}^{-1})^*\, \mathbf{P}_n \qquad (3)$$

i.e., $\tilde{\mathbf{A}}^{-1} \in \mathbb{CH}^{2n\times 2n}$. It is verified that, if $\tilde{\mathbf{A}} \in \mathbb{CH}^{2n\times 2m}$ and $\tilde{\mathbf{B}} \in \mathbb{CH}^{2m\times 2p}$, then $\tilde{\mathbf{A}}\, \tilde{\mathbf{B}} \in \mathbb{CH}^{2n\times 2p}$.

Centrohermitian 2-by-2 blockwise matrices can be mapped to matrices with real entries. Indeed, let us define the matrix

$$\mathbf{Q}_n = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_n & j\,\mathbf{I}_n \\ \mathbf{I}_n & -j\,\mathbf{I}_n \end{bmatrix} \in \mathbb{C}^{2n\times 2n} \qquad (4)$$

which is unitary (i.e., $\mathbf{Q}_n\, \mathbf{Q}_n^{\mathrm{H}} = \mathbf{I}_{2n}$) and its columns are conjugate symmetric (i.e., $\mathbf{P}_n\, \mathbf{Q}_n^* = \mathbf{Q}_n$). The mapping

$$\tilde{\mathbf{A}} \to \mathbf{Q}_n^{\mathrm{H}}\, \tilde{\mathbf{A}}\, \mathbf{Q}_m = \begin{bmatrix} \Re\{\mathbf{A}_{11} + \mathbf{A}_{12}^*\} & -\Im\{\mathbf{A}_{11} + \mathbf{A}_{12}^*\} \\ \Im\{\mathbf{A}_{11} - \mathbf{A}_{12}^*\} & \Re\{\mathbf{A}_{11} - \mathbf{A}_{12}^*\} \end{bmatrix} \qquad (5)$$

is bijective: it maps the subset $\mathbb{CH}^{2n\times 2m}$ onto the set of all the *real* 2-by-2 block matrices of the same size.

The mapping (5) between centrohermitian and real matrices allows to derive a useful property of the EVD of $\tilde{\mathbf{A}}$ given by (1) when it is square and Hermitian, i.e., $n = m$, $\mathbf{A}_{11}^{\mathrm{H}} = \mathbf{A}_{11}$, and $\mathbf{A}_{12} = \mathbf{A}_{12}^{\mathrm{T}}$. Let $\mathbf{Q}_n^{\mathrm{H}}\, \tilde{\mathbf{A}}\, \mathbf{Q}_n = \mathbf{W}\, \tilde{\boldsymbol{\Sigma}}\, \mathbf{W}^{\mathrm{T}}$ be the EVD of the real matrix $\mathbf{Q}_n^{\mathrm{H}}\, \tilde{\mathbf{A}}\, \mathbf{Q}_n$, where $\mathbf{W} \in \mathbb{R}^{2n\times 2n}$ is orthogonal (i.e., $\mathbf{W}\, \mathbf{W}^{\mathrm{T}} = \mathbf{I}_{2n}$) and $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{2n\times 2n}$ is diagonal.[2] Since $\mathbf{Q}_n$ is unitary, it follows that an EVD of $\tilde{\mathbf{A}}$ can be expressed as[3]

$$\tilde{\mathbf{A}} = (\mathbf{Q}_n \mathbf{W})\, \tilde{\boldsymbol{\Sigma}}\, (\mathbf{W}^{\mathrm{T}} \mathbf{Q}_n^{\mathrm{H}}) = \tilde{\mathbf{W}}\, \tilde{\boldsymbol{\Sigma}} \tilde{\mathbf{W}}^{\mathrm{H}} \qquad (6)$$

[2]The eigenvalues and eigenvectors of a (real) symmetric matrix are real.

[3]The EVD of a complex matrix is unique up to a unitary diagonal scaling matrix [17], provided that all the eigenvalues are distinct.

where it results that the matrix $\tilde{\mathbf{W}} \triangleq \mathbf{Q}_n\, \mathbf{W} \in \mathbb{C}^{2n\times 2n}$ is unitary (i.e., $\mathbf{Q}_n \mathbf{W}\, \mathbf{W}^{\mathrm{T}}\, \mathbf{Q}_n^{\mathrm{H}} = \mathbf{I}_{2n}$) and its columns are conjugate symmetric [i.e., $\mathbf{P}_n\, (\mathbf{Q}_n\, \mathbf{W})^* = \mathbf{P}_n\, \mathbf{Q}_n^*\, \mathbf{W} = \mathbf{Q}_n \mathbf{W}$]. The subset of all $2n \times k$ matrices having conjugate symmetric columns is denoted as $\mathbb{CS}^{2n\times k} \subset \mathbb{C}^{2n\times k}$, with $\mathbb{CS}^{2n} \subset \mathbb{C}^{2n}$ used as a shorthand for $\mathbb{CS}^{2n\times 1}$.

III. MIMO MODEL WITH WL PRECODING AND DECODING

The baseband-equivalent signal model describing transmission through a MIMO communication channel with N_{T} transmit and N_{R} receive antennas is given [1] by

$$\mathbf{y} = \mathbf{H}\, \mathbf{x} + \mathbf{n} \qquad (7)$$

where $\mathbf{y} \in \mathbb{C}^{N_{\mathrm{R}}}$ is the vector gathering the samples of the received signal, $\mathbf{x} \in \mathbb{C}^{N_{\mathrm{T}}}$ represents the transmitted precoded block, the channel matrix $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$ collects the coefficients of the underlying channels, which are assumed to be frequency-flat and constant during the transmission of the block $\mathbf{x}$, and $\mathbf{n} \in \mathbb{C}^{N_{\mathrm{R}}}$ denotes the zero-mean disturbance (i.e., interference plus noise) vector, with correlation matrix $\mathbf{R}_{\mathbf{nn}} \triangleq \mathbb{E}[\mathbf{n}\, \mathbf{n}^{\mathrm{H}}] \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{R}}}$ and (symmetric) conjugate correlation matrix $\mathbf{R}_{\mathbf{nn}^*} \triangleq \mathbb{E}[\mathbf{n}\, \mathbf{n}^{\mathrm{T}}] \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{R}}}$. The disturbance vector is proper iff $\mathbf{R}_{\mathbf{nn}^*} = \mathbf{O}_{N_{\mathrm{R}}\times N_{\mathrm{R}}}$, otherwise it is improper [7]. In what follows, we assume that: **a1)** $\mathbf{R}_{\mathbf{nn}}$ is positive definite (p.d.), i.e., $\det(\mathbf{R}_{\mathbf{nn}}) > 0$. If the channels are frequency selective, orthogonal frequency-division multiplexing techniques can be used to transform a frequency-selective MIMO channel into parallel frequency-non-selective MIMO channels, without loss of optimality [6]. Moreover, the forthcoming design is valid for any MIMO system and can be applied to other communication channels such as, e.g., frequency-selective channels employing transmit and receive filterbanks [2], [3], [5]; in this scenario, only the dimensionality of the vector/matrix variables in (7) and the structure of $\mathbf{H}$ will change.

In conventional precoding schemes [2]–[6], the transmitted vector $\mathbf{x}$ is a linear transformation of a zero-mean data vector $\mathbf{s} \in \mathbb{C}^{N_{\mathrm{S}}}$, which contains the N_{S} symbols to be transmitted, with $N_{\mathrm{S}} \leq R \triangleq \operatorname{rank}(\mathbf{H}) \leq \min(N_{\mathrm{T}}, N_{\mathrm{R}})$.[4] We propose to use a more general precoding rule (see Fig. 1), whereby the transmitted vector $\mathbf{x}$ is a WL transformation of $\mathbf{s}$:

$$\mathbf{x} = \mathbf{F}_1\, \mathbf{s} + \mathbf{F}_2\, \mathbf{s}^* \qquad (8)$$

[4]For linear precoding/decoding schemes, such a condition ensures perfect symbol recovery at the receiver in the absence of noise.

i.e., $\mathbf{x}$ jointly depends on $\mathbf{s}$ and its conjugate version $\mathbf{s}^*$, where $\mathbf{F}_1 \in \mathbb{C}^{N_\mathrm{T} \times N_\mathrm{S}}$ and $\mathbf{F}_2 \in \mathbb{C}^{N_\mathrm{T} \times N_\mathrm{S}}$ are *precoding matrices*. The maximum number of symbols per block will be implicitly optimized along with the explicit optimization of the WL transceiver. The average power per transmission is given by

$$\mathbb{E}[\|\mathbf{x}\|^2] = \mathrm{tr}(\mathbf{F}_1 \, \mathbf{R}_{\mathbf{ss}} \, \mathbf{F}_1^\mathrm{H}) + 2 \, \Re \left\{ \mathrm{tr}(\mathbf{F}_1 \, \mathbf{R}_{\mathbf{ss}^*} \, \mathbf{F}_2^\mathrm{H}) \right\} + \mathrm{tr}(\mathbf{F}_2 \, \mathbf{R}_{\mathbf{ss}}^* \, \mathbf{F}_2^\mathrm{H}) \quad (9)$$

where $\mathbf{R}_{\mathbf{ss}} \triangleq \mathbb{E}[\mathbf{s} \, \mathbf{s}^\mathrm{H}] \in \mathbb{C}^{N_\mathrm{S} \times N_\mathrm{S}}$ is the correlation matrix of $\mathbf{s}$ and $\mathbf{R}_{\mathbf{ss}^*} \triangleq \mathbb{E}[\mathbf{s} \, \mathbf{s}^\mathrm{T}] \in \mathbb{C}^{N_\mathrm{S} \times N_\mathrm{S}}$ denotes its (symmetric) conjugate correlation matrix. The symbol vector is proper iff $\mathbf{R}_{\mathbf{ss}^*} = \mathbf{O}_{N_\mathrm{S} \times N_\mathrm{S}}$, otherwise it is improper [7]. Hereinafter, we assume that: **a2)** $\mathbf{R}_{\mathbf{ss}}$ is p.d., i.e., $\det(\mathbf{R}_{\mathbf{ss}}) > 0$. Since the transmitter uses as a precoder a WL mapping of $\mathbf{s}$, defined by (8), it is quite natural to assume that the receiver performs an appropriate inverse WL mapping on the received vector $\mathbf{y}$ (see Fig. 1), thus yielding a WL estimate [11] $\mathbf{z} \in \mathbb{C}^{N_\mathrm{S}}$ of $\mathbf{s}$, which can be expressed as

$$\mathbf{z} = \mathbf{G}_1 \, \mathbf{y} + \mathbf{G}_2 \, \mathbf{y}^* = \mathbf{G} \, \tilde{\mathbf{y}} \quad (10)$$

where $\mathbf{G}_1 \in \mathbb{C}^{N_\mathrm{S} \times N_\mathrm{R}}$ and $\mathbf{G}_2 \in \mathbb{C}^{N_\mathrm{S} \times N_\mathrm{R}}$ are *decoding matrices*, $\mathbf{G} \triangleq [\mathbf{G}_1, \mathbf{G}_2] \in \mathbb{C}^{N_\mathrm{S} \times 2N_\mathrm{R}}$, and $\tilde{\mathbf{y}} \triangleq [\mathbf{y}^\mathrm{T}, \mathbf{y}^\mathrm{H}]^\mathrm{T} \in \mathbb{CS}^{2N_\mathrm{R}}$. Taking into account (7), the *augmented* complex vector is

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}} \, \mathbf{F} \, \mathbf{s} + \tilde{\mathbf{H}} \, \mathbf{P}_{N_\mathrm{T}} \, \mathbf{F}^* \mathbf{s}^* + \tilde{\mathbf{n}} \quad (11)$$

where

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \mathbf{H} & \mathbf{O}_{N_\mathrm{R} \times N_\mathrm{T}} \\ \mathbf{O}_{N_\mathrm{R} \times N_\mathrm{T}} & \mathbf{H}^* \end{bmatrix} \in \mathbb{CH}^{2N_\mathrm{R} \times 2N_\mathrm{T}} \quad (12)$$

$\mathbf{F} \triangleq [\mathbf{F}_1^T, \mathbf{F}_2^H]^T \in \mathbb{C}^{2N_\mathrm{T} \times N_\mathrm{S}}$, and $\tilde{\mathbf{n}} \triangleq [\mathbf{n}^\mathrm{T}, \mathbf{n}^\mathrm{H}]^\mathrm{T} \in \mathbb{CS}^{2N_\mathrm{R}}$. The correlation matrix of $\tilde{\mathbf{n}}$ can be expressed as

$$\mathbf{R}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}} \triangleq \mathbb{E}[\tilde{\mathbf{n}} \, \tilde{\mathbf{n}}^\mathrm{H}] = \begin{bmatrix} \mathbf{R}_{\mathbf{nn}} & \mathbf{R}_{\mathbf{nn}^*} \\ \mathbf{R}_{\mathbf{nn}^*}^* & \mathbf{R}_{\mathbf{nn}}^* \end{bmatrix} \in \mathbb{CH}^{2N_\mathrm{R} \times 2N_\mathrm{R}} . \quad (13)$$

It is noteworthy that, by setting $\mathbf{F}_2 = \mathbf{O}_{N_\mathrm{T} \times N_\mathrm{S}}$ in (8) and $\mathbf{G}_2 = \mathbf{O}_{N_\mathrm{S} \times N_\mathrm{R}}$ in (10), one obtains the conventional complex linear precoding/decoding scheme [2]–[6].

IV. Design of WL transceiver

A reasonable criterion to design the WL precoder $\mathbf{F}$ and the WL decoder $\mathbf{G}$, for a given $\mathbf{H}$, consists of minimizing the mean square error (MSE),[5] which is given by

$$\mathrm{MSE}(\mathbf{F}, \mathbf{G}) \triangleq \mathbb{E}[\|\mathbf{z} - \mathbf{s}\|^2] = \mathrm{tr}[\mathbf{E}(\mathbf{F}, \mathbf{G})] \quad (14)$$

where $\mathbf{E}(\mathbf{F}, \mathbf{G}) \triangleq \mathbb{E}[(\mathbf{z} - \mathbf{s}) \, (\mathbf{z} - \mathbf{s})^\mathrm{H}] \in \mathbb{C}^{N_\mathrm{S} \times N_\mathrm{S}}$ is the MSE matrix. As is always the case in practice, there is a power constraint on the precoding process, that is, $\mathbb{E}[\|\mathbf{x}\|^2] \le P_\mathrm{t}$, where $P_\mathrm{t} > 0$ is an application-specified bound on the transmitting power. Therefore, our goal is to design $\mathbf{F}$ (satisfying the aforementioned power constraint) and $\mathbf{G}$ such that $\mathrm{MSE}(\mathbf{F}, \mathbf{G})$ is minimized. In other words, we aim to solve the following constrained optimization problem

$$\min_{\mathbf{F}, \mathbf{G}} \mathrm{MSE}(\mathbf{F}, \mathbf{G}) \quad \text{subject to } \mathbb{E}[\|\mathbf{x}\|^2] \le P_\mathrm{t} \quad (15)$$

where $\mathbb{E}[\|\mathbf{x}\|^2]$ is given by (9). The following theorem (whose proof is omitted) simplifies the search for the solution of (15).

Theorem 1: Let

$$\tilde{\mathbf{F}} \triangleq [\mathbf{F}, \mathbf{P}_{N_\mathrm{T}} \, \mathbf{F}^*] = \begin{bmatrix} \mathbf{F}_1 & \mathbf{F}_2 \\ \mathbf{F}_2^* & \mathbf{F}_1^* \end{bmatrix} \in \mathbb{CH}^{2N_\mathrm{T} \times 2N_\mathrm{S}} \quad (16)$$

$$\tilde{\mathbf{G}} \triangleq [\mathbf{G}^\mathrm{T}, \mathbf{P}_{N_\mathrm{R}} \, \mathbf{G}^\mathrm{H}]^\mathrm{T} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 \\ \mathbf{G}_2^* & \mathbf{G}_1^* \end{bmatrix} \in \mathbb{CH}^{2N_\mathrm{S} \times 2N_\mathrm{R}} \quad (17)$$

and denote with

$$\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}} \triangleq \mathbb{E}[\tilde{\mathbf{s}} \, \tilde{\mathbf{s}}^\mathrm{H}] = \begin{bmatrix} \mathbf{R}_{\mathbf{ss}} & \mathbf{R}_{\mathbf{ss}^*} \\ \mathbf{R}_{\mathbf{ss}^*}^* & \mathbf{R}_{\mathbf{ss}}^* \end{bmatrix} \in \mathbb{CH}^{2N_\mathrm{S} \times 2N_\mathrm{S}} \quad (18)$$

the correlation matrix of $\tilde{\mathbf{s}} \triangleq [\mathbf{s}^\mathrm{T}, \mathbf{s}^\mathrm{H}]^\mathrm{T} \in \mathbb{CS}^{2N_\mathrm{S}}$. The matrices $\mathbf{F}_\mathrm{opt}$ and $\mathbf{G}_\mathrm{opt}$ are solutions of the original problem (15) iff the matrices $\tilde{\mathbf{F}}_\mathrm{opt} \triangleq [\mathbf{F}_\mathrm{opt}, \mathbf{P}_{N_\mathrm{T}} \, \mathbf{F}_\mathrm{opt}^*]$ and $\tilde{\mathbf{G}}_\mathrm{opt} \triangleq [\mathbf{G}_\mathrm{opt}^\mathrm{T}, \mathbf{P}_{N_\mathrm{R}} \, \mathbf{G}_\mathrm{opt}^\mathrm{H}]^\mathrm{T}$ are solutions of the following constrained minimization problem

$$\min_{\tilde{\mathbf{F}}, \tilde{\mathbf{G}}} \widetilde{\mathrm{MSE}}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}) \quad \text{subject to } \mathcal{P}(\tilde{\mathbf{F}}) \le 2 \, P_\mathrm{t} \quad (19)$$

with

$$\widetilde{\mathrm{MSE}}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}) \triangleq \mathbb{E}[\|\tilde{\mathbf{z}} - \tilde{\mathbf{s}}\|^2] = \mathrm{tr}[\tilde{\mathbf{E}}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}})] \quad (20)$$

$$\mathcal{P}(\tilde{\mathbf{F}}) \triangleq \mathrm{tr}(\tilde{\mathbf{F}} \, \mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}} \, \tilde{\mathbf{F}}^\mathrm{H}) \quad (21)$$

where $\tilde{\mathbf{E}}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}) \triangleq \mathbb{E}[(\tilde{\mathbf{z}} - \tilde{\mathbf{s}}) \, (\tilde{\mathbf{z}} - \tilde{\mathbf{s}})^\mathrm{H}] \in \mathbb{CH}^{2N_\mathrm{S} \times 2N_\mathrm{S}}$ is the *augmented* MSE matrix and $\tilde{\mathbf{z}} \triangleq [\mathbf{z}^\mathrm{T}, \mathbf{z}^\mathrm{H}]^\mathrm{T} \in \mathbb{CS}^{2N_\mathrm{S}}$.

Theorem 1 allows one to recast the original optimization problem (15) in the equivalent one (19), which can be solved by generalizing the mathematical framework used in [2], [4] and paying attention to the fact that $\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ might be singular. If the symbol vector is proper (i.e., $\mathbf{R}_{\mathbf{ss}^*} = \mathbf{O}_{N_\mathrm{S} \times N_\mathrm{S}}$), which necessarily requires that the symbols are complex-valued, e.g., they are drawn from a PSK or QAM constellation, then $\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ in (18) becomes block diagonal and, by virtue of assumption a2), it is nonsingular [i.e., $\mathrm{rank}(\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}) = \mathrm{rank}(\mathbf{R}_{\mathbf{ss}}) + \mathrm{rank}(\mathbf{R}_{\mathbf{ss}}^*) = 2N_\mathrm{S}$]. On the other hand, if improper modulation formats are used, such as ASK, differential BPSK (DBPSK), offset QPSK (OQPSK), offset QAM (OQAM), MSK and its variant Gaussian MSK (GMSK), it results [10], [12] that there is a linear deterministic relationship between $\mathbf{s}^*$ and $\mathbf{s}$, given by $\mathbf{s}^* = \mathbf{T} \, \mathbf{s}$, where $\mathbf{T} \in \mathbb{C}^{N_\mathrm{S} \times N_\mathrm{S}}$ is a diagonal unitary matrix (i.e., $\mathbf{T} \, \mathbf{T}^\mathrm{H} = \mathbf{I}_{N_\mathrm{S}}$).[6] In this case, it is easily seen that $\mathbf{R}_{\mathbf{ss}^*} = \mathbf{R}_{\mathbf{ss}} \, \mathbf{T}^*$ and, hence, $\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ is singular with $\mathrm{rank}(\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}) = \mathrm{rank}(\mathbf{R}_{\mathbf{ss}}) = N_\mathrm{S}$, since $\mathbf{R}_{\mathbf{ss}}/\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}} = \mathbf{R}_{\mathbf{ss}}^*(\mathbf{I}_{N_\mathrm{S}} - \mathbf{T} \, \mathbf{T}^*) = \mathbf{O}_{N_\mathrm{S} \times N_\mathrm{S}}$. Thus, we assume that $\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ is positive semidefinite (p.s.d.), which is equivalent to impose, in addition to $\det(\mathbf{R}_{\mathbf{ss}}) > 0$ [assumption a2)], that: **a3)** $\mathbf{R}_{\mathbf{ss}}/\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ is p.s.d., i.e., $\det(\mathbf{R}_{\mathbf{ss}}/\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}) \ge 0$. According to assumption a3), the eigenvalue decomposition (EVD) of $\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ can be written as

$$\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}} = \tilde{\mathbf{U}} \, \tilde{\mathbf{\Delta}} \, \tilde{\mathbf{U}}^\mathrm{H} \quad (22)$$

where $\tilde{\mathbf{\Delta}} \triangleq \mathrm{diag}[\tilde{\mathbf{\Delta}}_\mathrm{up}, \mathbf{O}_{(2N_\mathrm{S} - \tilde{L}) \times (2N_\mathrm{S} - \tilde{L})}] \in \mathbb{R}^{2N_\mathrm{S} \times 2N_\mathrm{S}}$, with $\tilde{\mathbf{\Delta}}_\mathrm{up} \triangleq \mathrm{diag}(\tilde{\delta}_1, \tilde{\delta}_2, \ldots, \tilde{\delta}_{\tilde{L}}) \in \mathbb{R}^{\tilde{L} \times \tilde{L}}$ gathering all the

[5] The proposed WL transceiver design can be generalized to the case of the weighted MMSE criterion [4] with minor modifications.

[6] In the case of real-valued symbols, it results that $\mathbf{T} = \mathbf{I}_{N_\mathrm{S}}$.

nonzero eigenvalues of $\mathbf{R}_{\tilde{s}\tilde{s}}$ and $N_S \le \tilde{L} \triangleq \text{rank}(\mathbf{R}_{\tilde{s}\tilde{s}}) \le 2\,N_S$, whereas $\tilde{\mathbf{U}} = [\tilde{\mathbf{U}}_{\text{left}}, \tilde{\mathbf{U}}_{\text{right}}] \in \mathbb{CS}^{2N_S \times 2N_S}$ collects the eigenvectors of $\mathbf{R}_{\tilde{s}\tilde{s}}$, with $\tilde{\mathbf{U}}_{\text{left}} \in \mathbb{CS}^{2N_S \times \tilde{L}}$, $\tilde{\mathbf{U}}_{\text{right}} \in \mathbb{CS}^{2N_S \times (2N_S - \tilde{L})}$, and $\tilde{\mathbf{U}}\,\tilde{\mathbf{U}}^{\text{H}} = \mathbf{I}_{2N_S}$. Since $\mathbf{R}_{\tilde{s}\tilde{s}} \in \mathbb{CH}^{2N_S \times 2N_S}$, its eigenvectors have been assumed to be conjugate symmetric (i.e., $\mathbf{P}_{N_S}\,\tilde{\mathbf{U}}^* = \tilde{\mathbf{U}}$).

Regarding the matrix $\mathbf{R}_{\tilde{n}\tilde{n}}$, we instead assume that it is p.d., which is equivalent to impose, in addition to $\det(\mathbf{R}_{\mathbf{nn}}) > 0$ [assumption a1)], that: **a4)** $\mathbf{R}_{\mathbf{nn}}/\mathbf{R}_{\tilde{n}\tilde{n}}$ is p.d., i.e., $\det(\mathbf{R}_{\mathbf{nn}}/\mathbf{R}_{\tilde{n}\tilde{n}}) > 0$. This assumption is justified by the fact that, in many cases of interest, an improper interference signal corrupts the MIMO transmission, in addition to proper thermal noise $\mathbf{w}$, i.e.,

$$\mathbf{n} = \mathbf{w} + \mathbf{H}_I\,\boldsymbol{\jmath} \tag{23}$$

where $\boldsymbol{\jmath} \in \mathbb{C}^{N_I}$ is independent of $\mathbf{w}$ and $\mathbf{H}_I \in \mathbb{C}^{N_R \times N_I}$ collects the channel coefficients between the interference source and the receiver: in this case, the correlation matrix of the disturbance is $\mathbf{R}_{\mathbf{nn}} = \mathbf{R}_{\mathbf{ww}} + \mathbf{H}_I\,\mathbf{R}_{\boldsymbol{\jmath\jmath}}\,\mathbf{H}_I^{\text{H}}$, where $\mathbf{R}_{\mathbf{ww}} \triangleq \mathbb{E}[\mathbf{w}\,\mathbf{w}^{\text{H}}] \in \mathbb{C}^{N_R \times N_R}$ is the correlation matrix of thermal noise and $\mathbf{R}_{\boldsymbol{\jmath\jmath}} \triangleq \mathbb{E}[\boldsymbol{\jmath}\,\boldsymbol{\jmath}^{\text{H}}] \in \mathbb{C}^{N_I \times N_I}$ is the correlation matrix of the interference, which are typically p.d., whereas its conjugate correlation matrix is $\mathbf{R}_{\mathbf{nn}^*} = \mathbf{H}_I\,\mathbf{R}_{\boldsymbol{\jmath\jmath}^*}\,\mathbf{H}_I^{\text{T}}$. Hence, the matrix $\mathbf{R}_{\mathbf{nn}}/\mathbf{R}_{\tilde{n}\tilde{n}}$ is typically nonsingular even when there is a linear deterministic relationship between $\boldsymbol{\jmath}^*$ and $\boldsymbol{\jmath}$.

Under the customary assumption that: **a5)** $\mathbb{E}[\mathbf{s}\,\mathbf{n}^{\text{T}}] = \mathbb{E}[\mathbf{s}\,\mathbf{n}^{\text{H}}] = \mathbf{O}_{N_S \times N_R}$, we observe that

$$\tilde{\mathbf{E}}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}) = \tilde{\mathbf{G}}\,\mathbf{R}_{\tilde{y}\tilde{y}}\,\tilde{\mathbf{G}}^{\text{H}} - \tilde{\mathbf{G}}\,\mathbf{R}_{\tilde{y}\tilde{s}} - \mathbf{R}_{\tilde{y}\tilde{s}}^{\text{H}}\,\tilde{\mathbf{G}}^{\text{H}} + \mathbf{R}_{\tilde{s}\tilde{s}} \tag{24}$$

where $\mathbf{R}_{\tilde{y}\tilde{y}} \triangleq \mathbb{E}[\tilde{\mathbf{y}}\,\tilde{\mathbf{y}}^{\text{H}}] \in \mathbb{CH}^{2N_R \times 2N_R}$ is the correlation matrix of $\tilde{\mathbf{y}}$ given by $\mathbf{R}_{\tilde{y}\tilde{y}} = \tilde{\mathbf{H}}\,\tilde{\mathbf{F}}\,\mathbf{R}_{\tilde{s}\tilde{s}}\,\tilde{\mathbf{F}}^{\text{H}}\,\tilde{\mathbf{H}}^{\text{H}} + \mathbf{R}_{\tilde{n}\tilde{n}}$, whereas $\mathbf{R}_{\tilde{y}\tilde{s}} \triangleq \mathbb{E}[\tilde{\mathbf{y}}\,\tilde{\mathbf{s}}^{\text{H}}] \in \mathbb{CH}^{2N_R \times N_S}$ is the cross-correlation matrix between $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{s}}$, which is given by $\mathbf{R}_{\tilde{y}\tilde{s}} = \tilde{\mathbf{H}}\,\tilde{\mathbf{F}}\,\mathbf{R}_{\tilde{s}\tilde{s}}$. The design problem (19) can be solved by resorting to the method of Lagrange duality. In this respect, let

$$\mathcal{L}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \mu) \triangleq \tilde{\text{MSE}}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}) + \mu\,[\mathcal{P}(\tilde{\mathbf{F}}) - 2\,P_t] \tag{25}$$

be the *Lagrangian*, where μ denotes the Lagrange multiplier, the potential solutions of (19) are the stationary points of $\mathcal{L}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \mu)$, that is, they satisfy the matrix equations

$$\nabla_{\tilde{\mathbf{G}}^*}\left[\mathcal{L}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \mu)\right] = \tilde{\mathbf{G}}\,\mathbf{R}_{\tilde{y}\tilde{y}} - \mathbf{R}_{\tilde{y}\tilde{s}}^{\text{H}} = \mathbf{O}_{2N_S \times 2N_R} \tag{26}$$

$$\nabla_{\tilde{\mathbf{F}}^*}\left[\mathcal{L}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \mu)\right] = \tilde{\mathbf{H}}^{\text{H}}\tilde{\mathbf{G}}^{\text{H}}\tilde{\mathbf{G}}\,\tilde{\mathbf{H}}\,\tilde{\mathbf{F}}\,\mathbf{R}_{\tilde{s}\tilde{s}} - \tilde{\mathbf{H}}^{\text{H}}\tilde{\mathbf{G}}^{\text{H}}\mathbf{R}_{\tilde{s}\tilde{s}} + \mu\,\tilde{\mathbf{F}}\,\mathbf{R}_{\tilde{s}\tilde{s}} = \mathbf{O}_{2N_T \times 2N_S} \tag{27}$$

with $\mu \ge 0$ and $\mu\,[\mathcal{P}(\tilde{\mathbf{F}}) - 2\,P_t] = 0$. It comes from (26) that the matrix $\tilde{\mathbf{G}}$ minimizing $\text{MSE}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}})$, given $\mathbf{H}$ and $\tilde{\mathbf{F}}$, can be explicated as

$$\tilde{\mathbf{G}} = \mathbf{R}_{\tilde{y}\tilde{s}}^{H}\,\mathbf{R}_{\tilde{y}\tilde{y}}^{-1} = \mathbf{R}_{\tilde{s}\tilde{s}}\,(\mathbf{I}_{2N_S} + \tilde{\mathbf{F}}^{\text{H}}\,\tilde{\mathbf{H}}^{\text{H}}\,\mathbf{R}_{\tilde{n}\tilde{n}}^{-1}\,\tilde{\mathbf{H}}\,\tilde{\mathbf{F}}\,\mathbf{R}_{\tilde{s}\tilde{s}})^{-1} \times \tilde{\mathbf{F}}^{\text{H}}\,\tilde{\mathbf{H}}^{\text{H}}\,\mathbf{R}_{\tilde{n}\tilde{n}}^{-1} \tag{28}$$

where we have also applied the matrix inversion lemma [17]. At this point, we deal with the difficult part of the transceiver design, which is the derivation of the optimum precoding $\tilde{\mathbf{F}}$. To simplify (26) and (27), one can resort to the following EVD

$$\tilde{\mathbf{H}}^{\text{H}}\,\mathbf{R}_{\tilde{n}\tilde{n}}^{-1}\,\tilde{\mathbf{H}} = \tilde{\mathbf{V}}\,\tilde{\boldsymbol{\Lambda}}\,\tilde{\mathbf{V}}^{\text{H}} \tag{29}$$

where $\tilde{\boldsymbol{\Lambda}} \triangleq \text{diag}[\tilde{\boldsymbol{\Lambda}}_{\text{up}}, \mathbf{O}_{(2N_T - \tilde{R}) \times (2N_T - \tilde{R})}] \in \mathbb{R}^{2N_T \times 2N_T}$, with $\tilde{\boldsymbol{\Lambda}}_{\text{up}} \triangleq \text{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_{\tilde{R}}) \in \mathbb{R}^{\tilde{R} \times \tilde{R}}$ gathering all the nonzero eigenvalues of $\tilde{\mathbf{H}}^{\text{H}}\,\mathbf{R}_{\tilde{n}\tilde{n}}^{-1}\,\tilde{\mathbf{H}}$ and $\tilde{R} \triangleq \text{rank}(\tilde{\mathbf{H}}^{\text{H}}\,\mathbf{R}_{\tilde{n}\tilde{n}}^{-1}\,\tilde{\mathbf{H}}) = \text{rank}(\tilde{\mathbf{H}}) \le \min(2\,N_R, 2\,N_T)$, whereas $\tilde{\mathbf{V}} = [\tilde{\mathbf{V}}_{\text{left}}, \tilde{\mathbf{V}}_{\text{right}}] \in \mathbb{CS}^{2N_T \times 2N_T}$ collects the eigenvectors of $\tilde{\mathbf{H}}^{\text{H}}\,\mathbf{R}_{\tilde{n}\tilde{n}}^{-1}\,\tilde{\mathbf{H}}$, with $\tilde{\mathbf{V}}_{\text{left}} \in \mathbb{CS}^{2N_T \times \tilde{R}}$, $\tilde{\mathbf{V}}_{\text{right}} \in \mathbb{CS}^{2N_T \times (2N_T - \tilde{R})}$, and $\tilde{\mathbf{V}}\,\tilde{\mathbf{V}}^{\text{H}} = \mathbf{I}_{2N_T}$. Since $\tilde{\mathbf{H}}^{\text{H}}\,\mathbf{R}_{\tilde{n}\tilde{n}}^{-1}\,\tilde{\mathbf{H}} \in \mathbb{CH}^{2N_T \times 2N_T}$, its eigenvectors have been assumed to be conjugate symmetric ($\mathbf{P}_{N_T}\,\tilde{\mathbf{V}}^* = \tilde{\mathbf{V}}$). By using the orthogonal decomposition theorem [17] and noticing that $\tilde{\mathbf{U}}$ is nonsingular, the matrix $\tilde{\mathbf{F}}$ can be equivalently decomposed as

$$\tilde{\mathbf{F}} = \tilde{\mathbf{V}}_{\text{left}}\,\tilde{\boldsymbol{\Phi}}_{\text{left}}\,\tilde{\mathbf{U}}^{\text{H}} + \tilde{\mathbf{V}}_{\text{right}}\,\tilde{\boldsymbol{\Phi}}_{\text{right}}\,\tilde{\mathbf{U}}^{\text{H}} \tag{30}$$

where $\tilde{\boldsymbol{\Phi}}_{\text{left}} \in \mathbb{R}^{\tilde{R} \times 2N_S}$ and $\tilde{\boldsymbol{\Phi}}_{\text{right}} \in \mathbb{R}^{(2N_T - \tilde{R}) \times 2N_S}$ are matrices to be determined. By substituting (30) in (28), it can be verified that, without any loss of generality, the optimum matrix $\tilde{\mathbf{G}}$ can be assumed to have the structure

$$\tilde{\mathbf{G}} = \tilde{\mathbf{U}}\,\tilde{\boldsymbol{\Psi}}\,\tilde{\boldsymbol{\Lambda}}_{\text{up}}^{-1}\,\tilde{\mathbf{V}}_{\text{left}}^{\text{H}}\,\tilde{\mathbf{H}}^{\text{H}}\mathbf{R}_{\tilde{n}\tilde{n}}^{-1} \tag{31}$$

where $\tilde{\boldsymbol{\Psi}} \in \mathbb{R}^{2N_S \times \tilde{R}}$ is a matrix to be determined. It is interesting to observe that, differently from the linear case, the matrices $\tilde{\boldsymbol{\Phi}}_{\text{left}}$, $\tilde{\boldsymbol{\Phi}}_{\text{right}}$, and $\tilde{\boldsymbol{\Psi}}$ must be necessarily real to ensure the centrohermitian 2-by-2 blockwise structure of $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{G}}$ (i.e., $\mathbf{P}_{N_T}\,\tilde{\mathbf{F}}^*\,\mathbf{P}_{N_S} = \tilde{\mathbf{F}}$ and $\mathbf{P}_{N_S}\,\tilde{\mathbf{G}}^*\,\mathbf{P}_{N_R} = \tilde{\mathbf{G}}$).

Without loss of generality, we assume that the eigenvalues of $\mathbf{R}_{\tilde{s}\tilde{s}}$ and $\tilde{\mathbf{H}}^{\text{H}}\,\mathbf{R}_{\tilde{n}\tilde{n}}^{-1}\,\tilde{\mathbf{H}}$ are arranged in decreasing order, i.e., $\tilde{\delta}_1 \ge \tilde{\delta}_2 \ge \cdots \ge \tilde{\delta}_{\tilde{L}} > 0$ and $\tilde{\lambda}_1 \ge \tilde{\lambda}_2 \ge \cdots \ge \tilde{\lambda}_{\tilde{R}} > 0$. The following theorem establishes the WL design algorithm.

Theorem 2: Assume that $\tilde{L} = \text{rank}(\mathbf{R}_{\tilde{s}\tilde{s}}) \le \tilde{R} = \text{rank}(\tilde{\mathbf{H}})$. The matrix $\tilde{\boldsymbol{\Phi}}_{\text{right}}$ does not affect the solution of (19) in anyway and, then, one can set $\tilde{\boldsymbol{\Phi}}_{\text{right}} = \mathbf{O}_{(2N_T - \tilde{R}) \times 2N_S}$, without loss of optimality, whereas the optimum matrices $\tilde{\boldsymbol{\Phi}}_{\text{left,opt}}$ and $\tilde{\boldsymbol{\Psi}}_{\text{opt}}$ are given by

$$\tilde{\boldsymbol{\Phi}}_{\text{left,opt}} = \tilde{\boldsymbol{\Theta}}_{\tilde{R}}\left(\mu^{-1/2}\,\tilde{\boldsymbol{\Delta}}_{\text{up}}^{-1/2}\,\tilde{\boldsymbol{\Lambda}}_{\text{up,red}}^{-1/2} - \tilde{\boldsymbol{\Delta}}_{\text{up}}^{-1}\,\tilde{\boldsymbol{\Lambda}}_{\text{up,red}}^{-1}\right)_+^{1/2}\tilde{\boldsymbol{\Theta}}_{N_S}^{T} \tag{32}$$

$$\tilde{\boldsymbol{\Psi}}_{\text{opt}} = \tilde{\boldsymbol{\Theta}}_{N_S}\left(\mu^{1/2}\,\tilde{\boldsymbol{\Delta}}_{\text{up}}^{1/2}\,\tilde{\boldsymbol{\Lambda}}_{\text{up,red}}^{-1/2} - \mu\,\tilde{\boldsymbol{\Lambda}}_{\text{up,red}}^{-1}\right)_+^{1/2}\tilde{\boldsymbol{\Theta}}_{\tilde{R}}^{\text{T}}\,\tilde{\boldsymbol{\Lambda}}_{\text{up}}^{1/2} \tag{33}$$

where $\tilde{\boldsymbol{\Theta}}_{\tilde{R}} \triangleq [\mathbf{I}_{\tilde{L}}, \mathbf{O}_{\tilde{L} \times (\tilde{R} - \tilde{L})}]^T \in \mathbb{R}^{\tilde{R} \times \tilde{L}}$, $\tilde{\boldsymbol{\Theta}}_{N_S} \triangleq [\mathbf{I}_{\tilde{L}}, \mathbf{O}_{\tilde{L} \times (2N_S - \tilde{L})}]^T \in \mathbb{R}^{2N_S \times \tilde{L}}$, and $\tilde{\boldsymbol{\Lambda}}_{\text{up,red}} \triangleq \text{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_{\tilde{L}}) \in \mathbb{R}^{\tilde{L} \times \tilde{L}}$, and for $K \le \tilde{L}$

$$\mu = \left(\frac{\sum_{k=1}^{K} \tilde{\delta}_k^{1/2}\,\tilde{\lambda}_k^{-1/2}}{2\,P_t + \sum_{k=1}^{K} \tilde{\lambda}_k^{-1}}\right)^2 . \tag{34}$$

Proof: The proof can be conducted by resorting to a mathematical framework similar to that used in [2], [4], by taking into account the (possible) singularity of $\mathbf{R}_{\tilde{s}\tilde{s}}$. ■

To optimally compute $\tilde{\boldsymbol{\Phi}}_{\text{left,opt}}$ and $\tilde{\boldsymbol{\Psi}}_{\text{opt}}$, one can resort to an iterative procedure similar to that used in the linear case [4], initialized with $K = \tilde{L}$. The WL precoding/decoding transceiver can potentially ensures a larger throughput with respect to linear structures.[7] When $\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ is nonsingular, i.e., $\tilde{L} = 2\,N_{\text{S}}$, which happens when the symbols are proper, e.g., PSK or QAM, condition $\tilde{L} \le \tilde{R}$ boils down to $N_{\text{S}} \le \min(N_{\text{T}}, N_{\text{R}})$ and, thus, the throughput of the WL transceiver is the same of its linear counterpart. However, when the symbols are ASK, DBPSK, OQPSK, OQAM, MSK, or GMSK, the matrix $\mathbf{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ is singular with $\tilde{L} = N_{\text{S}}$, in this case, condition $\tilde{L} \le \tilde{R}$ ends up to $N_{\text{S}} \le 2 \cdot \min(N_{\text{T}}, N_{\text{R}})$, which is exactly the double of the maximum number of symbols that can be transmitted by employing linear precoding/decoding techniques.

V. Numerical performance analysis

Results of Monte Carlo simulations are presented to assess the performance gain of the WL precoding/decoding technique (referred to as WL-Pre/Dec) over the corresponding linear counterpart of [2] (referred to as L-Pre/Dec). We considered an $(N_{\text{T}}, N_{\text{S}}, N_{\text{R}}, P_{\text{t}}) = (4, 4, 6, 1)$ MIMO system, transmitting QPSK symbols with $\mathbf{R}_{\mathbf{ss}} = \mathbf{I}_{N_{\text{S}}}$ (and $\mathbf{R}_{\mathbf{ss}^*} = \mathbf{O}_{N_{\text{S}} \times N_{\text{S}}}$) over a fading channel, where the entries of $\mathbf{H}$ are independent and identically distributed (i.i.d.) zero-mean unit-variance circularly symmetric complex Gaussian (CSCG) random variables.

The disturbance vector is modeled as in (23), where $\mathbf{w}$ is a zero-mean CSCG random vector with $\mathbf{R}_{\mathbf{ww}} = \sigma_w^2\,\mathbf{I}_{N_{\text{R}}}$, the elements of $\mathbf{H}_{\text{I}}$ are i.i.d. zero-mean unit-variance CSCG random variables, and $\boldsymbol{\jmath}$ is composed of BPSK symbols with $N_{\text{I}} = 6$ and $\mathbf{R}_{\boldsymbol{\jmath}\boldsymbol{\jmath}} = \mathbf{R}_{\boldsymbol{\jmath}\boldsymbol{\jmath}^*} = \mathbf{I}_{N_{\text{I}}}$. The signal-to-disturbance ratio (SDR) is defined as $\text{SDR} \triangleq \mathbb{E}[\|\mathbf{H}\,\mathbf{x}\|^2]/\mathbb{E}[\|\mathbf{n}\|^2]$, whereas $\text{INR} \triangleq \mathbb{E}[\|\mathbf{H}_{\text{I}}\,\boldsymbol{\jmath}\|^2]/\mathbb{E}[\|\mathbf{w}\|^2]$ is the interference-to-noise ratio (INR). The average (over channel realizations) bit-error rate (BER) versus the SDR is plotted in Fig. 1, for different disturbance scenarios: (i) the disturbance is proper (i.e., $\text{INR} = 0$); (ii) the disturbance is composed of proper and improper equal-power components (i.e., $\text{INR} = 1$); (ii) the disturbance is completely improper (i.e., $\text{INR} = +\infty$).

Results show that, when the disturbance is proper, the WL transceiver achieves the same BER performance of its linear counterpart, thus confirming that linear precoding/decoding schemes are optimal in the case of proper symbols and disturbance. However, when the disturbance contains an improper component, the WL transceiver significantly outperforms the linear one, by ensuring huge performance gains when the disturbance is completely improper.

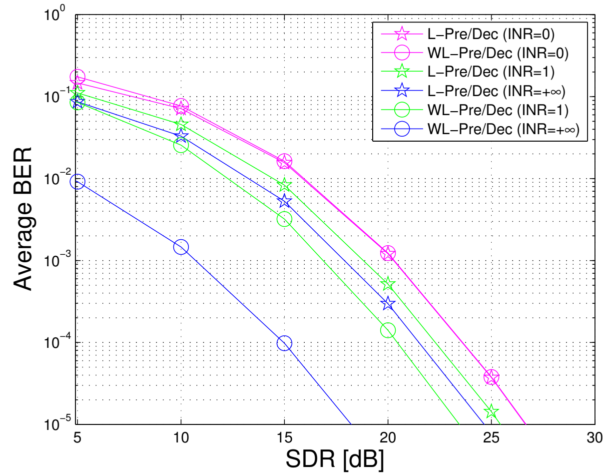

Figure 2. ABER versus SDR for different disturbance scenarios.

[7] This fact was also observed in [15] without, however, linking this property to the rank of the second-order statistics of the transmitted symbol block.

References

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[2] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmission over dispersive channels", *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1019–1032, Apr. 1999.

[3] A. Scaglione, G. B. Giannakis and S. Barbarossa, "Redundant filterbank precoders and equalizers Part I: unification and optimal designs," *IEEE Trans. Signal Process.*, vol. 47, pp. 1988–2006, July 1999.

[4] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, pp. 2198–2206, Dec. 2001.

[5] A. Scaglione, P. Stoica, S. Barbarossa, G.B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, pp. 1051–1064, May 2002.

[6] D.P. Palomar, J.M. Cioffi, and M.A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, pp. 2381–2401 , Sept. 2003.

[7] F.D. Neeser and J.L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Information Theory*, pp. 1293–1302, July 1993.

[8] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, pp. 3473–3482, Dec. 1994.

[9] P.J. Schreier and L.L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Processing*, pp. 714–725, Mar. 2003.

[10] D. Darsena, G. Gelli, L. Paura and F. Verde, "Subspace-based blind channel identification of SISO-FIR systems with improper random inputs", *Eurasip Journal on Signal Processing, Special Issue on Signal Processing in Communications*, vol. 84, pp. 2021–2039, Nov. 2004.

[11] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, pp. 2030–2033, Aug. 1995.

[12] W.H. Gerstacker, R. Schober and A. Lampe, "Receivers with widely linear processing for frequency-selective channels," *IEEE Trans. Commun.*, vol. 51, pp. 1512–1523, Sep. 2003.

[13] D. Darsena, G. Gelli, L. Paura and F. Verde, "Widely-linear equalization and blind channel identification for interference-contaminated multicarrier systems," *IEEE Trans. Signal Processing*, pp. 1163–1177, March 2005.

[14] D. Darsena, G. Gelli, and F. Verde, "Universal linear precoding for NBI-proof widely-linear equalization in MC systems," *Eurasip Journal on Wireless Communications and Networking*, vol. 2008, Article ID 321450, 13 pages.

[15] F. Sterle, "Widely linear MMSE transceiver for MIMO channels," *IEEE Trans. Signal Process.*, vol. 55, pp. 4258–4270, Aug. 2007.

[16] D. Darsena, G. Gelli, L. Paura, and F. Verde, " Widely-linear beamforming/combining techniques for MIMO wireless systems," in *Proc. of 5th International Symp. Commun. Control and Signal Proc.*, Rome, Italy, May 2012, pp. 1–5.

[17] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[18] A. Lee, "Centrohermitian and skew-centrohermitian matrices," *Linear Algebra and its Applications*, vol. 29, pp. 205-210, 1980.