Unified Theory of Complex-Valued Matrix Differentiation

Are Hjørungnes
UniK - University Graduate Center
University of Oslo
Instituttveien 25, P. O. Box 70
N-2027 Kjeller, Norway
Email: arehj@unik.no

David Gesbert
Mobile Communications Department
Eurécom Institute
2229 Route des Crêtes, BP 193
F-06904 Sophia Antipolis Cédex, France
Email: gesbert@eurecom.fr

Daniel P. Palomar
Department of Electronic and Computer Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong
Email: palomar@ust.hk

***Abstract*— A systematic theory is introduced for finding the derivatives of complex-valued matrix functions with respect to a complex-valued matrix variable and the complex conjugate of this variable. In the framework introduced, the differential of the complex-valued matrix function is used to identify the derivatives of this function. Matrix differentiation results are developed for use in signal processing and communications applications. Several other examples are given.**

Index Terms: Adaptive filters, Optimization methods, Gradient methods, Linear algebra

I. INTRODUCTION

In many engineering problems, the unknown parameters are complex-valued vectors and matrices and, often, the task of the system designer is to find the values of these complex parameters which optimize a chosen criterion. This problem is of particular interest in recent signal processing for communications applications where the optimization of complex matrices such as transmit MIMO precoders, receiver filters, equalizer, or transmit covariance matrix have become extremely common. Often the cost function represents a system performance metric such as the channel capacity, SNR, or error rate. When a scalar real-valued function depends on a complex-valued matrix parameter, the necessary conditions for optimality can of course be found by either setting the derivative of the function with respect to the complex-valued matrix parameter or its complex conjugate to a zero vector/matrix. Differentiation results are well-known for certain classes of functions, e.g., quadratic functions. However when the form taken by the cost-function is too intricate, the researcher is left with guess work, trying to arrive at an expression for the derivative.

This paper provides tools for finding derivatives in a systematic way that will help students and researchers alike. The tools can also be used to determine the direction of maximum rate of change of a real-valued scalar function, with respect to the complex-valued matrix parameter, in view of use in iterative algorithms. Our results offer a generalization of a well-known results for scalar functions of vector variables. The main contribution of this paper is to generalize the real-valued derivatives given in [1] to the complex-valued case. In particular, we propose to do so by finding the derivatives by the so-called complex differentials of the functions. In this paper, it is assumed that the functions are differentiable with respect to the complex-valued parameter matrix and its complex conjugate, and it will be seen that these two parameter matrices should be treated as independent when finding the derivatives, as is classical for scalar variables.

The problem at hand has been treated for *real-valued* matrix variables in [1], [2], [3], [4], [5]. Four additional references that give a brief treatment of the case of real-valued scalar functions which depend complex-valued vectors are Appendix B of [6], Appendix 2.B in [7] and the article [8]. The article [9] serves as an introduction to this area for complex-valued scalar functions with complex-valued argument vectors. Results on complex differentiation theory is given in [10], [11] for differentiation with respect to complex-valued scalars and vectors, however, the more general *matrix* case is not considered. In [12], they find derivatives of scalar functions with respect to complex-valued matrices, however, that paper could have been considerably simplified if the proposed theory was utilized. Examples of problems where the unknown matrix is a complex-valued matrix are wide ranging including precoding of MIMO systems [13], linear equalization design [14], array signal processing [15] to cite a few.

Some of the most relevant applications to signal and communication problems are presented here, with key results being highlighted and other illustrative examples being listed in tables. For an extended version, see [16], [17].

The rest of this paper is organized as follows: In Section II, the complex differential is introduced, and based on this differential, the definition of the derivatives of complex-valued matrix function with respect to the complex-valued matrix argument and its complex conjugate is given in Section III. The key procedure showing how the derivatives can be found from the differential of a function is also presented in Section III. Section IV contains the important results of equivalent conditions for finding stationary points and in which direction the function has the maximum rate of change. In Section V, several key results are placed in tables and some results are derived for various cases with high relevance for signal processing and communication problems. Section VII contains some conclusions.

Notation: Scalar quantities (variables z or functions f) are denoted by lowercase symbols, vector quantities (variables $\boldsymbol{z}$ or functions $\boldsymbol{f}$) are denoted by lowercase boldface symbols, and matrix quantities (variables $\boldsymbol{Z}$ or functions $\boldsymbol{F}$) are denoted by capital boldface symbols. It is assumed that all the functions depend on a complex variable $\boldsymbol{Z}$ and the complex conjugate of the same variable $\boldsymbol{Z}^*$. Furthermore, it is assumed that all the elements of $\boldsymbol{Z}$ are independent such that they can be freely chosen. Let $\jmath = \sqrt{-1}$, and let the real $\mathrm{Re}\{\cdot\}$ and imaginary $\mathrm{Im}\{\cdot\}$ operators return the real and imaginary parts of the input matrix, respectively. If $\boldsymbol{Z} \in \mathbb{C}^{N\times Q}$ is a complex-valued[1] matrix, then $\boldsymbol{Z} = \mathrm{Re}\{\boldsymbol{Z}\} + \jmath\,\mathrm{Im}\{\boldsymbol{Z}\}$, and $\boldsymbol{Z}^* = \mathrm{Re}\{\boldsymbol{Z}\} - \jmath\,\mathrm{Im}\{\boldsymbol{Z}\}$, where $\mathrm{Re}\{\boldsymbol{Z}\} \in \mathbb{R}^{N\times Q}$, $\mathrm{Im}\{\boldsymbol{Z}\} \in \mathbb{R}^{N\times Q}$, and the operator $(\cdot)^*$ denotes complex conjugate of the matrix it is applied to.

II. COMPLEX DIFFERENTIALS

The differential has the same size as the matrix it is applied to. The differential can be found component-wise, that is, $(d\boldsymbol{Z})_{k,l} = d(\boldsymbol{Z})_{k,l}$. A procedure that can often be used for finding the differentials of a complex-valued matrix function[2] $\boldsymbol{F}(\boldsymbol{Z}_0, \boldsymbol{Z}_1)$ is to calculate the difference

$$\begin{aligned}\boldsymbol{F}(\boldsymbol{Z}_0 + d\boldsymbol{Z}_0, \boldsymbol{Z}_1 + d\boldsymbol{Z}_1) - \boldsymbol{F}(\boldsymbol{Z}_0, \boldsymbol{Z}_1) &= \text{First-order}(d\boldsymbol{Z}_0, d\boldsymbol{Z}_1)\\ &+ \text{Higher-order}(d\boldsymbol{Z}_0, d\boldsymbol{Z}_1), \qquad (1)\end{aligned}$$

where First-order$(\cdot,\cdot)$ contains the terms that depend on either $d\boldsymbol{Z}_0$ or $d\boldsymbol{Z}_1$ of the first order, and Higher-order$(\cdot,\cdot)$ denotes

This work was supported by the Research Council of Norway (NFR) and the French Ministry of Foreign Affairs through the Aurora project entitled "Optimization of Broadband Wireless Communications Network" and NFR project 176773/S10.

[1] $\mathbb{R}$ and $\mathbb{C}$ are the sets of the real and complex numbers, respectively.
[2] The indexes are chosen to start with 0 everywhere in this article.

TABLE I

IMPORTANT RESULTS FOR COMPLEX DIFFERENTIALS.

$d\boldsymbol{A} = \boldsymbol{0}$	$d(a\boldsymbol{Z}) = a d\boldsymbol{Z}$	$d(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{B}) = \boldsymbol{A}(d\boldsymbol{Z})\boldsymbol{B}$	$d(\boldsymbol{Z}_0 + \boldsymbol{Z}_1) = d\boldsymbol{Z}_0 + d\boldsymbol{Z}_1$
$d \operatorname{Tr}\{\boldsymbol{Z}\} = \operatorname{Tr}\{d\boldsymbol{Z}\}$	$d(\boldsymbol{Z}_0\boldsymbol{Z}_1) = (d\boldsymbol{Z}_0)\boldsymbol{Z}_1 + \boldsymbol{Z}_0(d\boldsymbol{Z}_1)$	$d(\boldsymbol{Z}_0 \otimes \boldsymbol{Z}_1) = (d\boldsymbol{Z}_0) \otimes \boldsymbol{Z}_1 + \boldsymbol{Z}_0 \otimes (d\boldsymbol{Z}_1)$	$d \operatorname{Tr}\{\exp(\boldsymbol{Z})\} = \operatorname{Tr}\{\exp(\boldsymbol{Z})d\boldsymbol{Z}\}$
$d\boldsymbol{Z}^* = (d\boldsymbol{Z})^*$	$d\boldsymbol{Z}^H = (d\boldsymbol{Z})^H$	$d \det(\boldsymbol{Z}) = \det(\boldsymbol{Z}) \operatorname{Tr}\left\{\boldsymbol{Z}^{-1} d\boldsymbol{Z}\right\}$	$d \ln(\det(\boldsymbol{Z})) = \operatorname{Tr}\left\{\boldsymbol{Z}^{-1} d\boldsymbol{Z}\right\}$
$d \operatorname{reshape}(\boldsymbol{Z}) = \operatorname{reshape}(d\boldsymbol{Z})$	$d(\boldsymbol{Z}_0 \odot \boldsymbol{Z}_1) = (d\boldsymbol{Z}_0) \odot \boldsymbol{Z}_1 + \boldsymbol{Z}_0 \odot (d\boldsymbol{Z}_1)$	$d\boldsymbol{Z}^{-1} = -\boldsymbol{Z}^{-1}(d\boldsymbol{Z})\boldsymbol{Z}^{-1}$	$d \operatorname{Tr}\{\boldsymbol{Z}^n\} = n \operatorname{Tr}\{\boldsymbol{Z}^{n-1} d\boldsymbol{Z}\}$

the terms that depend on the higher order terms of $d\boldsymbol{Z}_0$ and $d\boldsymbol{Z}_1$. The differential is then given by First-order$(\cdot,\cdot)$, i.e., the first order term of $\boldsymbol{F}(\boldsymbol{Z}_0 + d\boldsymbol{Z}_0, \boldsymbol{Z}_1 + d\boldsymbol{Z}_1) - \boldsymbol{F}(\boldsymbol{Z}_0, \boldsymbol{Z}_1)$. Observe that $\lim_{(d\boldsymbol{Z}_0, d\boldsymbol{Z}_1)\to\boldsymbol{0}} \frac{\text{Higher-order}(d\boldsymbol{Z}_0, d\boldsymbol{Z}_1)}{\|(d\boldsymbol{Z}_0, d\boldsymbol{Z}_1)\|} = \boldsymbol{0}$. Therefore, for sufficiently small $(d\boldsymbol{Z}_0, d\boldsymbol{Z}_1)$, we can write the first order (affine) approximation $\boldsymbol{F}(\boldsymbol{Z}_0 + d\boldsymbol{Z}_0, \boldsymbol{Z}_1 + d\boldsymbol{Z}_1) \approx \boldsymbol{F}(\boldsymbol{Z}_0, \boldsymbol{Z}_1) +$ First-order$(d\boldsymbol{Z}_0, d\boldsymbol{Z}_1)$. As an example, let $\boldsymbol{F}(\boldsymbol{Z}_0, \boldsymbol{Z}_1) = \boldsymbol{Z}_0\boldsymbol{Z}_1$. Then the difference in (1) can be developed and readily expressed as: $\boldsymbol{F}(\boldsymbol{Z}_0 + d\boldsymbol{Z}_0, \boldsymbol{Z}_1 + d\boldsymbol{Z}_1) - \boldsymbol{F}(\boldsymbol{Z}_0, \boldsymbol{Z}_1) = \boldsymbol{Z}_0 d\boldsymbol{Z}_1 + (d\boldsymbol{Z}_0)\boldsymbol{Z}_1 + (d\boldsymbol{Z}_0)(d\boldsymbol{Z}_1)$. The differential of $\boldsymbol{Z}_0\boldsymbol{Z}_1$ can then be identified as all the first-order terms on either $d\boldsymbol{Z}_0$ or $d\boldsymbol{Z}_1$ as $d\boldsymbol{Z}_0\boldsymbol{Z}_1 = \boldsymbol{Z}_0 d\boldsymbol{Z}_1 + (d\boldsymbol{Z}_0)\boldsymbol{Z}_1$.

Let $\otimes$ and $\odot$ denote the Kronecker and Hadamard product [18], respectively. Some important rules on complex differentials are listed in Table I, assuming $\boldsymbol{A}$, $\boldsymbol{B}$, and a to be constants, $n \in \{1, 2, 3, \ldots\}$, and $\boldsymbol{Z}$, $\boldsymbol{Z}_0$, and $\boldsymbol{Z}_1$ to be complex-valued matrix variables. The vectorization operator $\operatorname{vec}(\cdot)$ stacks the columns vectors of the argument matrix into a long column vector in chronological order [18]. The differentiation rule of the reshaping operator $\operatorname{reshape}(\cdot)$ in Table I is valid for any *linear reshaping*[3] operator $\operatorname{reshape}(\cdot)$ of the matrix, and examples of such operators are the transpose $(\cdot)^T$ or $\operatorname{vec}(\cdot)$. Some of the basic differential results in Table I can be derived by means of (1), and others can be derived by generalizing some of the results found in [1], [4] to the complex differential case. From Table I, the following four equalities follows $d\boldsymbol{Z} = d\operatorname{Re}\{\boldsymbol{Z}\} + \jmath d\operatorname{Im}\{\boldsymbol{Z}\}$, $d\boldsymbol{Z}^* = d\operatorname{Re}\{\boldsymbol{Z}\} - \jmath d\operatorname{Im}\{\boldsymbol{Z}\}$, $d\operatorname{Re}\{\boldsymbol{Z}\} = \frac{1}{2}(d\boldsymbol{Z} + d\boldsymbol{Z}^*)$, and $d\operatorname{Im}\{\boldsymbol{Z}\} = \frac{1}{2\jmath}(d\boldsymbol{Z} - d\boldsymbol{Z}^*)$.

The following lemma is used to identify the first-order derivatives later in the article. The real variables $\operatorname{Re}\{\boldsymbol{Z}\}$ and $\operatorname{Im}\{\boldsymbol{Z}\}$ are independent of each other and hence are their differentials. Although the complex variables $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$ are related, their differentials are linearly independent in the following way:

Lemma 1: *Let $\boldsymbol{Z} \in \mathbb{C}^{N\times Q}$ and let $\boldsymbol{A}_i \in \mathbb{C}^{M\times NQ}$. If $\boldsymbol{A}_0 d\operatorname{vec}(\boldsymbol{Z}) + \boldsymbol{A}_1 d\operatorname{vec}(\boldsymbol{Z}^*) = \boldsymbol{0}_{M\times 1}$ for all $d\boldsymbol{Z} \in \mathbb{C}^{N\times Q}$, then $\boldsymbol{A}_i = \boldsymbol{0}_{M\times NQ}$ for $i \in \{0, 1\}$.*

Proof: *See [17].* ■

III. COMPUTATION OF THE DERIVATIVE WITH RESPECT TO COMPLEX-VALUED MATRICES

The most general definition of the derivative is given here from which the definitions for less general cases follow. In this article, it is assumed that all the elements of the matrix $\boldsymbol{Z}$ are linearly independent.

Definition 1 (Derivatives by Differential): Let $\boldsymbol{F} : \mathbb{C}^{N\times Q} \times \mathbb{C}^{N\times Q} \to \mathbb{C}^{M\times P}$. Then the derivative of the matrix function $\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*) \in \mathbb{C}^{M\times P}$ with respect to $\boldsymbol{Z} \in \mathbb{C}^{N\times Q}$ is denoted $\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F}$, and the derivative of the matrix function $\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*) \in \mathbb{C}^{M\times P}$ with respect to $\boldsymbol{Z}^* \in \mathbb{C}^{N\times Q}$ is denoted $\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F}$ and the size of both these derivatives is $MP \times NQ$. The derivatives $\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F}$ and $\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F}$ are defined by the following differential expression:

$$d\operatorname{vec}(\boldsymbol{F}) = (\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F})\, d\operatorname{vec}(\boldsymbol{Z}) + (\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F})\, d\operatorname{vec}(\boldsymbol{Z}^*). \tag{2}$$

$\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*)$ and $\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*)$ are called the *Jacobian* matrices of $\boldsymbol{F}$ with respect to $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$, respectively.

Remark 1: *Definition* 1 *is a generalization of Definition 1 in [1, p. 173] to include complex-valued matrices. In [1], several alternative definitions of the derivative of real-valued functions with respect to a matrix are discussed, and it is concluded that the definition that matches Definition* 1 *is the only reasonable definition. Definition* 1 *is also a generalization of the definition used in [9] for complex-valued* vectors *to the case of complex-valued* matrices*.*

Assume that $d\operatorname{vec}(\boldsymbol{F}) = \boldsymbol{\zeta}_0 d\operatorname{vec}(\boldsymbol{Z}) + \boldsymbol{\zeta}_1 d\operatorname{vec}(\boldsymbol{Z}^*)$ where $\boldsymbol{\zeta}_i \in \mathbb{C}^{MP\times NQ}$, and $\boldsymbol{\zeta}_1$ might be a function $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$. To show the uniqueness of the representation in (2), we subtract the differential in (2) from $d\operatorname{vec}(\boldsymbol{F}) = \boldsymbol{\zeta}_0 d\operatorname{vec}(\boldsymbol{Z}) + \boldsymbol{\zeta}_1 d\operatorname{vec}(\boldsymbol{Z}^*)$ to get $(\boldsymbol{\zeta}_0 - \mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*))\, d\operatorname{vec}(\boldsymbol{Z}) + (\boldsymbol{\zeta}_1 - \mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*))\, d\operatorname{vec}(\boldsymbol{Z}^*) = \boldsymbol{0}_{MP\times 1}$. Using Lemma 1, it follows that the derivative is unique.

Definition 2 (Partial Derivatives): Let $\boldsymbol{f} : \mathbb{C}^{N\times 1} \times \mathbb{C}^{N\times 1} \to \mathbb{C}^{M\times 1}$. The partial derivatives $\frac{\partial}{\partial \boldsymbol{z}^T}\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{z}^*)$ and $\frac{\partial}{\partial \boldsymbol{z}^H}\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{z}^*)$ [4] of size $M \times N$ are defined as

$$\left(\frac{\partial}{\partial \boldsymbol{z}^T}\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{z}^*)\right)_{k,l} = \frac{\partial}{\partial z_k} f_l, \tag{3}$$

$$\left(\frac{\partial}{\partial \boldsymbol{z}^H}\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{z}^*)\right)_{k,l} = \frac{\partial}{\partial z_k^*} f_l, \tag{4}$$

where z_i and f_i is component number i of the vectors $\boldsymbol{z}$ and $\boldsymbol{f}$, respectively.

Notice that $\frac{\partial}{\partial \boldsymbol{z}^T}\boldsymbol{f} = \mathcal{D}_{\boldsymbol{z}}\boldsymbol{f}$ and $\frac{\partial}{\partial \boldsymbol{z}^H}\boldsymbol{f} = \mathcal{D}_{\boldsymbol{z}^*}\boldsymbol{f}$. Using the partial derivative notation in Definition 2, the derivatives of the function $\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*)$, in Definition 1, are:

$$\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*) = \frac{\partial \operatorname{vec}(\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*))}{\partial \operatorname{vec}^T(\boldsymbol{Z})}, \tag{5}$$

$$\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*) = \frac{\partial \operatorname{vec}(\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*))}{\partial \operatorname{vec}^T(\boldsymbol{Z}^*)}. \tag{6}$$

This is a generalization of the real-valued matrix variable case treated in [1] to the complex-valued matrix variable case. (5) and (6) show how the all the $MPNQ$ partial derivatives of all the components of $\boldsymbol{F}$ with respect to all the components of $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$ are arranged when using the notation introduced in Definition 2.

Key result: Finding the derivative of the complex-valued matrix function $\boldsymbol{F}$ with respect to the complex-valued matrices $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$ can be achieved using the following three-step procedure:

1) *Compute the differential* $d\operatorname{vec}(\boldsymbol{F})$*.*
2) *Manipulate the expression into the form given* (2)*.*
3) *Read out the result using Definition* 1*.*

For less general function types, a similar procedure can be used.

Chain Rule: One big advantage of the way the derivative is defined in Definition 1 compared to other definitions of the derivative of $\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*)$ is that the chain rule is valid in a very simple form. The chain rule is now formulated, and it might be very useful for finding complicated derivatives.

Theorem 1 (Chain Rule with Differentials): Let $(S_0, S_1) \subseteq \mathbb{C}^{N\times Q} \times \mathbb{C}^{N\times Q}$, and let $\boldsymbol{F} : S_0 \times S_1 \to \mathbb{C}^{M\times P}$ be differentiable with respect to both its first and second argument at an interior point $(\boldsymbol{Z}, \boldsymbol{Z}^*)$ in the set $S_0 \times S_1$. Let $T_0 \times T_1 \subseteq \mathbb{C}^{M\times P} \times \mathbb{C}^{M\times P}$

[3] The output of the reshape operator has the same number of elements as the input, but the shape of the output might be different, so $\operatorname{reshape}(\cdot)$ performs a reshaping of its input argument.

[4] In this article, $(\cdot)^H$ denotes the complex conjugate transpose.

TABLE II

DERIVATIVES OF SCALAR-VALUED FUNCTIONS $f\left(\boldsymbol{Z}, \boldsymbol{Z}^*\right) \in \mathbb{C}$

$f\left(\boldsymbol{Z}, \boldsymbol{Z}^*\right)$	Differential df	$\frac{\partial}{\partial \boldsymbol{Z}} f$	$\frac{\partial}{\partial \boldsymbol{Z}^*} f$
$\mathrm{Tr}\{\boldsymbol{Z}\}$	$\mathrm{Tr}\left\{\boldsymbol{I}_N d\boldsymbol{Z}\right\}$	$\boldsymbol{I}_N$	$\boldsymbol{0}_{N\times N}$
$\mathrm{Tr}\{\boldsymbol{Z}^*\}$	$\mathrm{Tr}\left\{\boldsymbol{I}_N d\boldsymbol{Z}^*\right\}$	$\boldsymbol{0}_{N\times N}$	$\boldsymbol{I}_N$
$\mathrm{Tr}\{\boldsymbol{A}\boldsymbol{Z}\}$	$\mathrm{Tr}\left\{\boldsymbol{A} d\boldsymbol{Z}\right\}$	$\boldsymbol{A}^T$	$\boldsymbol{0}_{N\times Q}$
$\mathrm{Tr}\{\boldsymbol{Z}^H\boldsymbol{A}\}$	$\mathrm{Tr}\left\{\boldsymbol{A}^T d\boldsymbol{Z}^*\right\}$	$\boldsymbol{0}_{N\times Q}$	$\boldsymbol{A}$
$\mathrm{Tr}\{\boldsymbol{Z}\boldsymbol{A}_0\boldsymbol{Z}^T\boldsymbol{A}_1\}$	$\mathrm{Tr}\left\{\left(\boldsymbol{A}_0\boldsymbol{Z}^T\boldsymbol{A}_1 + \boldsymbol{A}_0^T\boldsymbol{Z}^T\boldsymbol{A}_1^T\right) d\boldsymbol{Z}\right\}$	$\boldsymbol{A}_1^T\boldsymbol{Z}\boldsymbol{A}_0^T + \boldsymbol{A}_1\boldsymbol{Z}\boldsymbol{A}_0$	$\boldsymbol{0}_{N\times Q}$
$\mathrm{Tr}\{\boldsymbol{Z}\boldsymbol{A}_0\boldsymbol{Z}\boldsymbol{A}_1\}$	$\mathrm{Tr}\left\{\left(\boldsymbol{A}_0\boldsymbol{Z}\boldsymbol{A}_1 + \boldsymbol{A}_1\boldsymbol{Z}\boldsymbol{A}_0\right) d\boldsymbol{Z}\right\}$	$\boldsymbol{A}_1^T\boldsymbol{Z}^T\boldsymbol{A}_0^T + \boldsymbol{A}_0^T\boldsymbol{Z}^T\boldsymbol{A}_1^T$	$\boldsymbol{0}_{N\times Q}$
$\mathrm{Tr}\{\boldsymbol{Z}\boldsymbol{A}_0\boldsymbol{Z}^H\boldsymbol{A}_1\}$	$\mathrm{Tr}\left\{\boldsymbol{A}_0\boldsymbol{Z}^H\boldsymbol{A}_1 d\boldsymbol{Z} + \boldsymbol{A}_0^T\boldsymbol{Z}^T\boldsymbol{A}_1^T d\boldsymbol{Z}^*\right\}$	$\boldsymbol{A}_1^T\boldsymbol{Z}^*\boldsymbol{A}_0^T$	$\boldsymbol{A}_1\boldsymbol{Z}\boldsymbol{A}_0$
$\mathrm{Tr}\{\boldsymbol{Z}\boldsymbol{A}_0\boldsymbol{Z}^*\boldsymbol{A}_1\}$	$\mathrm{Tr}\left\{\boldsymbol{A}_0\boldsymbol{Z}^*\boldsymbol{A}_1 d\boldsymbol{Z} + \boldsymbol{A}_1\boldsymbol{Z}\boldsymbol{A}_0 d\boldsymbol{Z}^*\right\}$	$\boldsymbol{A}_1^T\boldsymbol{Z}^H\boldsymbol{A}_0^T$	$\boldsymbol{A}_0^T\boldsymbol{Z}^T\boldsymbol{A}_1^T$
$\mathrm{Tr}\{\boldsymbol{A}\boldsymbol{Z}^{-1}\}$	$-\mathrm{Tr}\left\{\boldsymbol{Z}^{-1}\boldsymbol{A}\boldsymbol{Z}^{-1} d\boldsymbol{Z}\right\}$	$-\left(\boldsymbol{Z}^T\right)^{-1}\boldsymbol{A}^T\left(\boldsymbol{Z}^T\right)^{-1}$	$\boldsymbol{0}_{N\times N}$
$\mathrm{Tr}\{\boldsymbol{Z}^p\}$	$p\,\mathrm{Tr}\left\{\boldsymbol{Z}^{p-1} d\boldsymbol{Z}\right\}$	$p\left(\boldsymbol{Z}^T\right)^{p-1}$	$\boldsymbol{0}_{N\times N}$
$\ln\left(\det(\boldsymbol{A}_0\boldsymbol{Z}\boldsymbol{A}_1)\right)$	$\mathrm{Tr}\left\{\boldsymbol{A}_1\left(\boldsymbol{A}_0\boldsymbol{Z}\boldsymbol{A}_1\right)^{-1}\boldsymbol{A}_0 d\boldsymbol{Z}\right\}$	$\boldsymbol{A}_0^T\left(\boldsymbol{A}_1^T\boldsymbol{Z}^T\boldsymbol{A}_0^T\right)^{-1}\boldsymbol{A}_1^T$	$\boldsymbol{0}_{N\times Q}$
$\ln\left(\det(\boldsymbol{Z}\boldsymbol{Z}^T)\right)$	$2\,\mathrm{Tr}\left\{\boldsymbol{Z}^T\left(\boldsymbol{Z}\boldsymbol{Z}^T\right)^{-1} d\boldsymbol{Z}\right\}$	$2\left(\boldsymbol{Z}\boldsymbol{Z}^T\right)^{-1}\boldsymbol{Z}$	$\boldsymbol{0}_{N\times Q}$
$\ln\left(\det(\boldsymbol{Z}\boldsymbol{Z}^*)\right)$	$\mathrm{Tr}\left\{\boldsymbol{Z}^*(\boldsymbol{Z}\boldsymbol{Z}^*)^{-1} d\boldsymbol{Z} + (\boldsymbol{Z}\boldsymbol{Z}^*)^{-1}\boldsymbol{Z} d\boldsymbol{Z}^*\right\}$	$(\boldsymbol{Z}^H\boldsymbol{Z}^T)^{-1}\boldsymbol{Z}^H$	$\boldsymbol{Z}^T\left(\boldsymbol{Z}^H\boldsymbol{Z}^T\right)^{-1}$
$\ln\left(\det(\boldsymbol{Z}\boldsymbol{Z}^H)\right)$	$\mathrm{Tr}\left\{\boldsymbol{Z}^H(\boldsymbol{Z}\boldsymbol{Z}^H)^{-1} d\boldsymbol{Z} + \boldsymbol{Z}^T\left(\boldsymbol{Z}\boldsymbol{Z}^H\right)^{-T} d\boldsymbol{Z}^*\right\}$	$(\boldsymbol{Z}\boldsymbol{Z}^H)^{-T}\boldsymbol{Z}^*$	$\left(\boldsymbol{Z}\boldsymbol{Z}^H\right)^{-1}\boldsymbol{Z}$
$\ln\left(\det(\boldsymbol{Z}^p)\right)$	$p\,\mathrm{Tr}\left\{\boldsymbol{Z}^{-1} d\boldsymbol{Z}\right\}$	$p\left(\boldsymbol{Z}^T\right)^{-1}$	$\boldsymbol{0}_{N\times N}$
$\ln\left(\det\left(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H\right)\right)$	$\mathrm{Tr}\left\{\boldsymbol{Z}^H\left(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H\right)^{-1} d\boldsymbol{Z} + \boldsymbol{Z}^T(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H)^{-T} d\boldsymbol{Z}^*\right\}$	$\left(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H\right)^{-T}\boldsymbol{Z}^*$	$(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H)^{-1}\boldsymbol{Z}$
$\lambda(\boldsymbol{Z})$	$\frac{\boldsymbol{v}_0^H(d\boldsymbol{Z})\boldsymbol{u}_0}{\boldsymbol{v}_0^H\boldsymbol{u}_0} = \mathrm{Tr}\left\{\frac{\boldsymbol{u}_0\boldsymbol{v}_0^H}{\boldsymbol{v}_0^H\boldsymbol{u}_0} d\boldsymbol{Z}\right\}$	$\frac{\boldsymbol{v}_0^*\boldsymbol{u}_0^T}{\boldsymbol{v}_0^H\boldsymbol{u}_0}$	$\boldsymbol{0}_{N\times N}$
$\lambda^*(\boldsymbol{Z})$	$\frac{\boldsymbol{v}_0^T(d\boldsymbol{Z}^*)\boldsymbol{u}_0^*}{\boldsymbol{v}_0^T\boldsymbol{u}_0^*} = \mathrm{Tr}\left\{\frac{\boldsymbol{u}_0^*\boldsymbol{v}_0^T}{\boldsymbol{v}_0^T\boldsymbol{u}_0^*} d\boldsymbol{Z}^*\right\}$	$\boldsymbol{0}_{N\times N}$	$\frac{\boldsymbol{v}_0\boldsymbol{u}_0^H}{\boldsymbol{v}_0^T\boldsymbol{u}_0^*}$

be such that $(\boldsymbol{F}(\boldsymbol{Z},\boldsymbol{Z}^*), \boldsymbol{F}^*(\boldsymbol{Z},\boldsymbol{Z}^*)) \in T_0 \times T_1$ for all $(\boldsymbol{Z},\boldsymbol{Z}^*) \in S_0 \times S_1$. Assume that $\boldsymbol{G} : T_0 \times T_1 \to \mathbb{C}^{R\times S}$ is differentiable at an interior point $(\boldsymbol{F}(\boldsymbol{Z},\boldsymbol{Z}^*), \boldsymbol{F}^*(\boldsymbol{Z},\boldsymbol{Z}^*)) \in T_0 \times T_1$. Define the composite function $\boldsymbol{H} : S_0 \times S_1 \to \mathbb{C}^{R\times S}$ by $\boldsymbol{H}\left(\boldsymbol{Z},\boldsymbol{Z}^*\right) \triangleq \boldsymbol{G}\left(\boldsymbol{F}(\boldsymbol{Z},\boldsymbol{Z}^*), \boldsymbol{F}^*(\boldsymbol{Z},\boldsymbol{Z}^*)\right)$. The derivatives $\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{H}$ and $\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{H}$ are obtained through:

$$\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{H} = \left(\mathcal{D}_{\boldsymbol{F}}\boldsymbol{G}\right)\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F} + \left(\mathcal{D}_{\boldsymbol{F}^*}\boldsymbol{G}\right)\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F}^*, \tag{7}$$

$$\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{H} = \left(\mathcal{D}_{\boldsymbol{F}}\boldsymbol{G}\right)\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F} + \left(\mathcal{D}_{\boldsymbol{F}^*}\boldsymbol{G}\right)\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F}^*. \tag{8}$$

Proof: *See [17].* ■

IV. COMPLEX DERIVATIVES IN OPTIMIZATION THEORY

In this section, two useful theorems are presented that exploit the theory introduced earlier. Both theorems are important when solving practical optimization problems involving differentiation with respect to a complex-valued matrix. These results include conditions for finding stationary points for a real-valued scalar function dependent on complex-valued matrices and in which direction the same type of function has the minimum or maximum rate of change, which might be used in the *steepest decent method*.

1) **Stationary Points:** The next theorem presents conditions for finding stationary points of $f(\boldsymbol{Z},\boldsymbol{Z}^*) \in \mathbb{R}$.

Theorem 2: *Let $f : \mathbb{C}^{N\times Q} \times \mathbb{C}^{N\times Q} \to \mathbb{R}$. A stationary point*[5] *of the function $f(\boldsymbol{Z},\boldsymbol{Z}^*) = g(\boldsymbol{X},\boldsymbol{Y})$, where $g : \mathbb{R}^{N\times Q} \times \mathbb{R}^{N\times Q} \to \mathbb{R}$ and $\boldsymbol{Z} = \boldsymbol{X} + \jmath\boldsymbol{Y}$ is then found by one of the following three equivalent conditions: (a) $\mathcal{D}_{\boldsymbol{X}}g(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{0}_{1\times NQ} \;\wedge\; \mathcal{D}_{\boldsymbol{Y}}g(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{0}_{1\times NQ}$, (b) $\mathcal{D}_{\boldsymbol{Z}}f(\boldsymbol{Z},\boldsymbol{Z}^*) = \boldsymbol{0}_{1\times NQ}$, or (c) $\mathcal{D}_{\boldsymbol{Z}^*}f(\boldsymbol{Z},\boldsymbol{Z}^*) = \boldsymbol{0}_{1\times NQ}$.*

Proof: *See [17].* ■

2) **Direction of Extremal Rate of Change:** The next theorem states how to find the maximum and minimum rate of change of $f(\boldsymbol{Z},\boldsymbol{Z}^*) \in \mathbb{R}$.

Theorem 3: *Let $f : \mathbb{C}^{N\times Q} \times \mathbb{C}^{N\times Q} \to \mathbb{R}$. The directions where the function f have the maximum and minimum rate of change with respect to $\mathrm{vec}(\boldsymbol{Z})$ are given by $\left[\mathcal{D}_{\boldsymbol{Z}^*}f(\boldsymbol{Z},\boldsymbol{Z}^*)\right]^T$ and $-\left[\mathcal{D}_{\boldsymbol{Z}^*}f(\boldsymbol{Z},\boldsymbol{Z}^*)\right]^T$, respectively.*

Proof: *See [17].* ■

[5]Notice that a stationary point can be a local minimum, a local maximum, or a saddle point.

V. LINKS WITH CLASSICAL GRADIENT

For scalar-valued functions $f\left(\boldsymbol{Z},\boldsymbol{Z}^*\right) \in \mathbb{C}$, it is common to arrange the partial derivatives $\frac{\partial}{\partial z_{k,l}} f$ and $\frac{\partial}{\partial z_{k,l}^*} f$ in an alternative way [1] than in the expressions $\mathcal{D}_{\boldsymbol{Z}} f\left(\boldsymbol{Z},\boldsymbol{Z}^*\right)$ and $\mathcal{D}_{\boldsymbol{Z}^*} f\left(\boldsymbol{Z},\boldsymbol{Z}^*\right)$. The notation for the alternative way of organizing all the partial derivatives is $\frac{\partial}{\partial \boldsymbol{Z}} f$ and $\frac{\partial}{\partial \boldsymbol{Z}^*} f$. In this alternative way, the partial derivatives of the elements of the matrix $\boldsymbol{Z} \in \mathbb{C}^{N\times Q}$ are arranged as:

$$\left(\frac{\partial}{\partial \boldsymbol{Z}} f\right)_{k,l} = \frac{\partial}{\partial z_{k,l}} f, \tag{9}$$

$$\left(\frac{\partial}{\partial \boldsymbol{Z}^*} f\right)_{k,l} = \frac{\partial}{\partial z_{k,l}^*} f. \tag{10}$$

$\frac{\partial}{\partial \boldsymbol{Z}} f$ and $\frac{\partial}{\partial \boldsymbol{Z}^*} f$ are called the *gradient*[6] of f with respect to $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$, respectively. (9) generalizes to the complex case of one of the ways to define the derivative of real-valued scalar functions with respect to real matrices in [1]. The way of arranging the partial derivatives in (9) is different than than the way given in (5) and (6). If $df = \mathrm{vec}^T(\boldsymbol{A}_0)d\,\mathrm{vec}(\boldsymbol{Z}) + \mathrm{vec}^T(\boldsymbol{A}_1)d\,\mathrm{vec}(\boldsymbol{Z}^*) = \mathrm{Tr}\left\{\boldsymbol{A}_0^T d\boldsymbol{Z} + \boldsymbol{A}_1^T d\boldsymbol{Z}^*\right\}$, where $\boldsymbol{A}_i, \boldsymbol{Z} \in \mathbb{C}^{N\times Q}$, then it can be shown that $\frac{\partial}{\partial \boldsymbol{Z}} f = \boldsymbol{A}_0$ and $\frac{\partial}{\partial \boldsymbol{Z}^*} f = \boldsymbol{A}_1$, where the matrices $\boldsymbol{A}_0$ and $\boldsymbol{A}_1$ depend on $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$ in general. The size of $\frac{\partial}{\partial \boldsymbol{Z}} f$ and $\frac{\partial}{\partial \boldsymbol{Z}^*} f$ is $N \times Q$, while the size of $\mathcal{D}_{\boldsymbol{Z}} f\left(\boldsymbol{Z},\boldsymbol{Z}^*\right)$ and $\mathcal{D}_{\boldsymbol{Z}^*} f\left(\boldsymbol{Z},\boldsymbol{Z}^*\right)$ is $1 \times NQ$, so these two ways of organizing the partial derivatives are different. It can be shown, that $\mathcal{D}_{\boldsymbol{Z}} f\left(\boldsymbol{Z},\boldsymbol{Z}^*\right) = \mathrm{vec}^T\left(\frac{\partial}{\partial \boldsymbol{Z}} f\left(\boldsymbol{Z},\boldsymbol{Z}^*\right)\right)$, and $\mathcal{D}_{\boldsymbol{Z}^*} f\left(\boldsymbol{Z},\boldsymbol{Z}^*\right) = \mathrm{vec}^T\left(\frac{\partial}{\partial \boldsymbol{Z}^*} f\left(\boldsymbol{Z},\boldsymbol{Z}^*\right)\right)$. The steepest decent method can be formulated as $\boldsymbol{Z}_{k+1} = \boldsymbol{Z}_k + \mu\frac{\partial}{\partial \boldsymbol{Z}^*} f(\boldsymbol{Z}_k, \boldsymbol{Z}_k^*)$.

VI. SPCOM APPLICATIONS AND EXAMPLES

Due to space limitations, we limit ourselves to developing two key examples of cost functions and derivatives in the text below, with particular interest in communications systems optimization. For other useful examples of cost functions, we simply apply the provided theory and summarize the results in Table II for scalar functions and in Table III for matrix functions.

A. Determinant Related Problems

Cost functions that depend on the determinant appear in many signal processing and communications related problems. Particularly so in the capacity of wireless multiple-input multiple-output (MIMO)

[6]The following notation also exists [6], [12] for the gradient $\nabla_{\boldsymbol{Z}} f \triangleq \frac{\partial}{\partial \boldsymbol{Z}^*} f$.

TABLE III

DERIVATIVES OF MATRIX-VALUED FUNCTIONS $\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*)$

$\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*)$	Differential $d\,\mathrm{vec}(\boldsymbol{F})$	$\mathcal{D}_{\boldsymbol{Z}}\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*)$	$\mathcal{D}_{\boldsymbol{Z}^*}\boldsymbol{F}(\boldsymbol{Z}, \boldsymbol{Z}^*)$
$\boldsymbol{Z}$	$\boldsymbol{I}_{NQ} d\,\mathrm{vec}(\boldsymbol{Z})$	$\boldsymbol{I}_{NQ}$	$\boldsymbol{0}_{NQ \times NQ}$
$\boldsymbol{Z}^T$	$\boldsymbol{K}_{N,Q} d\,\mathrm{vec}(\boldsymbol{Z})$	$\boldsymbol{K}_{N,Q}$	$\boldsymbol{0}_{NQ \times NQ}$
$\boldsymbol{Z}^*$	$\boldsymbol{I}_{NQ} d\,\mathrm{vec}(\boldsymbol{Z}^*)$	$\boldsymbol{0}_{NQ \times NQ}$	$\boldsymbol{I}_{NQ}$
$\boldsymbol{Z}^H$	$\boldsymbol{K}_{N,Q} d\,\mathrm{vec}(\boldsymbol{Z}^*)$	$\boldsymbol{0}_{NQ \times NQ}$	$\boldsymbol{K}_{N,Q}$
$\boldsymbol{Z}\boldsymbol{Z}^T$	$\left(\boldsymbol{I}_{N^2} + \boldsymbol{K}_{N,N}\right)(\boldsymbol{Z} \otimes \boldsymbol{I}_N)\, d\,\mathrm{vec}(\boldsymbol{Z})$	$\left(\boldsymbol{I}_{N^2} + \boldsymbol{K}_{N,N}\right)(\boldsymbol{Z} \otimes \boldsymbol{I}_N)$	$\boldsymbol{0}_{N^2 \times NQ}$
$\boldsymbol{Z}^T\boldsymbol{Z}$	$\left(\boldsymbol{I}_{Q^2} + \boldsymbol{K}_{Q,Q}\right)\left(\boldsymbol{I}_Q \otimes \boldsymbol{Z}^T\right) d\,\mathrm{vec}(\boldsymbol{Z})$	$\left(\boldsymbol{I}_{Q^2} + \boldsymbol{K}_{Q,Q}\right)\left(\boldsymbol{I}_Q \otimes \boldsymbol{Z}^T\right)$	$\boldsymbol{0}_{Q^2 \times NQ}$
$\boldsymbol{Z}\boldsymbol{Z}^H$	$(\boldsymbol{Z}^* \otimes \boldsymbol{I}_N)\, d\,\mathrm{vec}(\boldsymbol{Z}) + \boldsymbol{K}_{N,N}(\boldsymbol{Z} \otimes \boldsymbol{I}_N)\, d\,\mathrm{vec}(\boldsymbol{Z}^*)$	$\boldsymbol{Z}^* \otimes \boldsymbol{I}_N$	$\boldsymbol{K}_{N,N}(\boldsymbol{Z} \otimes \boldsymbol{I}_N)$
$\boldsymbol{Z}^{-1}$	$-\left((\boldsymbol{Z}^T)^{-1} \otimes \boldsymbol{Z}^{-1}\right) d\,\mathrm{vec}(\boldsymbol{Z})$	$-(\boldsymbol{Z}^T)^{-1} \otimes \boldsymbol{Z}^{-1}$	$\boldsymbol{0}_{N^2 \times N^2}$
$\boldsymbol{Z}^p$	$\sum_{i=1}^{p}\left((\boldsymbol{Z}^T)^{p-i} \otimes \boldsymbol{Z}^{i-1}\right) d\,\mathrm{vec}(\boldsymbol{Z})$	$\sum_{i=1}^{p}\left((\boldsymbol{Z}^T)^{p-i} \otimes \boldsymbol{Z}^{i-1}\right)$	$\boldsymbol{0}_{N^2 \times N^2}$
$\boldsymbol{Z} \otimes \boldsymbol{Z}$	$(\boldsymbol{A}(\boldsymbol{Z}) + \boldsymbol{B}(\boldsymbol{Z}))d\,\mathrm{vec}(\boldsymbol{Z})$	$\boldsymbol{A}(\boldsymbol{Z}) + \boldsymbol{B}(\boldsymbol{Z})$	$\boldsymbol{0}_{N^2Q^2 \times NQ}$
$\boldsymbol{Z} \otimes \boldsymbol{Z}^*$	$\boldsymbol{A}(\boldsymbol{Z}^*)d\,\mathrm{vec}(\boldsymbol{Z}) + \boldsymbol{B}(\boldsymbol{Z})d\,\mathrm{vec}(\boldsymbol{Z}^*)$	$\boldsymbol{A}(\boldsymbol{Z}^*)$	$\boldsymbol{B}(\boldsymbol{Z})$
$\boldsymbol{Z}^* \otimes \boldsymbol{Z}^*$	$(\boldsymbol{A}(\boldsymbol{Z}^*) + \boldsymbol{B}(\boldsymbol{Z}^*))d\,\mathrm{vec}(\boldsymbol{Z}^*)$	$\boldsymbol{0}_{N^2Q^2 \times NQ}$	$\boldsymbol{A}(\boldsymbol{Z}^*) + \boldsymbol{B}(\boldsymbol{Z}^*)$
$\boldsymbol{Z} \odot \boldsymbol{Z}$	$2\,\mathrm{diag}(\mathrm{vec}(\boldsymbol{Z}))d\,\mathrm{vec}(\boldsymbol{Z})$	$2\,\mathrm{diag}(\mathrm{vec}(\boldsymbol{Z}))$	$\boldsymbol{0}_{NQ \times NQ}$
$\boldsymbol{Z} \odot \boldsymbol{Z}^*$	$\mathrm{diag}(\mathrm{vec}(\boldsymbol{Z}^*))d\,\mathrm{vec}(\boldsymbol{Z}) + \mathrm{diag}(\mathrm{vec}(\boldsymbol{Z}))d\,\mathrm{vec}(\boldsymbol{Z}^*)$	$\mathrm{diag}(\mathrm{vec}(\boldsymbol{Z}^*))$	$\mathrm{diag}(\mathrm{vec}(\boldsymbol{Z}))$
$\boldsymbol{Z}^* \odot \boldsymbol{Z}^*$	$2\,\mathrm{diag}(\mathrm{vec}(\boldsymbol{Z}^*))d\,\mathrm{vec}(\boldsymbol{Z}^*)$	$\boldsymbol{0}_{NQ \times NQ}$	$2\,\mathrm{diag}(\mathrm{vec}(\boldsymbol{Z}^*))$

communication systems [14], as well as in upper bounds for the pairwise error probability (PEP) [13].

Let $f : \mathbb{C}^{N\times Q} \times \mathbb{C}^{Q\times N} \rightarrow \mathbb{C}$ be $f(\boldsymbol{Z}, \boldsymbol{Z}^*) = \ln\left(\det(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H)\right)$, where $\boldsymbol{P} \in \mathbb{C}^{N\times N}$ is independent of $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$. df is found by the rules in Table I as $df = \mathrm{Tr}\left\{(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H)^{-1} d(\boldsymbol{Z}\boldsymbol{Z}^H)\right\} = \mathrm{Tr}\left\{\boldsymbol{Z}^H\left(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H\right)^{-1} d\boldsymbol{Z} + \boldsymbol{Z}^T(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H)^{-T} d\boldsymbol{Z}^*\right\}$. From this, the derivatives with respect to $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$ of $\ln\left(\det\left(\boldsymbol{P} + \boldsymbol{Z}\boldsymbol{Z}^H\right)\right)$ can be found, and they are included in Table II.

B. Kronecker Product Related Problems

An objective functions which depends on the Kronecker product of the unknown complex-valued matrix is the PEP found in [13]. Let $\boldsymbol{K}_{N,Q}$ denote the *commutation matrix* [1], and let $\boldsymbol{F} : \mathbb{C}^{N_0\times Q_0} \times \mathbb{C}^{N_1\times Q_1} \rightarrow \mathbb{C}^{N_0N_1\times Q_0Q_1}$ be given by $\boldsymbol{F}(\boldsymbol{Z}_0, \boldsymbol{Z}_1) = \boldsymbol{Z}_0 \otimes \boldsymbol{Z}_1$, where $\boldsymbol{Z}_i \in \mathbb{C}^{N_i\times Q_i}$. The differential of this function follows from Table I: $d\boldsymbol{F} = (d\boldsymbol{Z}_0)\otimes\boldsymbol{Z}_1 + \boldsymbol{Z}_0\otimes d\boldsymbol{Z}_1$. Applying the $\mathrm{vec}(\cdot)$ operator to $d\boldsymbol{F}$ yields: $d\,\mathrm{vec}(\boldsymbol{F}) = \mathrm{vec}\left((d\boldsymbol{Z}_0)\otimes\boldsymbol{Z}_1\right) + \mathrm{vec}\left(\boldsymbol{Z}_0 \otimes d\boldsymbol{Z}_1\right)$. From Theorem 3.10 in [1], it follows that

$$\begin{aligned} \mathrm{vec}\left((d\boldsymbol{Z}_0)\otimes\boldsymbol{Z}_1\right) &= (\boldsymbol{I}_{Q_0}\otimes\boldsymbol{K}_{Q_1,N_0}\otimes\boldsymbol{I}_{N_1})\left[(d\,\mathrm{vec}(\boldsymbol{Z}_0))\otimes\mathrm{vec}(\boldsymbol{Z}_1)\right] \\ &= (\boldsymbol{I}_{Q_0}\otimes\boldsymbol{K}_{Q_1,N_0}\otimes\boldsymbol{I}_{N_1})\left[\boldsymbol{I}_{N_0Q_0}\otimes\mathrm{vec}(\boldsymbol{Z}_1)\right] d\,\mathrm{vec}(\boldsymbol{Z}_0), \quad (11) \end{aligned}$$

and in a similar way it follows that: $\mathrm{vec}\left(\boldsymbol{Z}_0 \otimes d\boldsymbol{Z}_1\right) = (\boldsymbol{I}_{Q_0}\otimes\boldsymbol{K}_{Q_1,N_0}\otimes\boldsymbol{I}_{N_1})\left[\mathrm{vec}(\boldsymbol{Z}_0)\otimes\boldsymbol{I}_{N_1Q_1}\right] d\,\mathrm{vec}(\boldsymbol{Z}_1)$. Inserting the last two results into $d\,\mathrm{vec}(\boldsymbol{F})$ gives:

$$\begin{aligned} d\,\mathrm{vec}(\boldsymbol{F}) &= (\boldsymbol{I}_{Q_0}\otimes\boldsymbol{K}_{Q_1,N_0}\otimes\boldsymbol{I}_{N_1})\left[\boldsymbol{I}_{N_0Q_0}\otimes\mathrm{vec}(\boldsymbol{Z}_1)\right] d\,\mathrm{vec}(\boldsymbol{Z}_0) \\ &+ (\boldsymbol{I}_{Q_0}\otimes\boldsymbol{K}_{Q_1,N_0}\otimes\boldsymbol{I}_{N_1})\left[\mathrm{vec}(\boldsymbol{Z}_0)\otimes\boldsymbol{I}_{N_1Q_1}\right] d\,\mathrm{vec}(\boldsymbol{Z}_1). \quad (12) \end{aligned}$$

Define the matrices $\boldsymbol{A}(\boldsymbol{Z}_1)$ and $\boldsymbol{B}(\boldsymbol{Z}_0)$ by $\boldsymbol{A}(\boldsymbol{Z}_1) \triangleq (\boldsymbol{I}_{Q_0}\otimes\boldsymbol{K}_{Q_1,N_0}\otimes\boldsymbol{I}_{N_1})\left[\boldsymbol{I}_{N_0Q_0}\otimes\mathrm{vec}(\boldsymbol{Z}_1)\right]$, and $\boldsymbol{B}(\boldsymbol{Z}_0) = (\boldsymbol{I}_{Q_0}\otimes\boldsymbol{K}_{Q_1,N_0}\otimes\boldsymbol{I}_{N_1})\left[\mathrm{vec}(\boldsymbol{Z}_0)\otimes\boldsymbol{I}_{N_1Q_1}\right]$. It is then possible to rewrite the differential of $\boldsymbol{F}(\boldsymbol{Z}_0, \boldsymbol{Z}_1) = \boldsymbol{Z}_0 \otimes \boldsymbol{Z}_1$ as $d\,\mathrm{vec}(\boldsymbol{F}) = \boldsymbol{A}(\boldsymbol{Z}_1)d\,\mathrm{vec}(\boldsymbol{Z}_0) + \boldsymbol{B}(\boldsymbol{Z}_0)d\,\mathrm{vec}(\boldsymbol{Z}_1)$. From $d\,\mathrm{vec}(\boldsymbol{F})$, the differentials and derivatives of $\boldsymbol{Z}\otimes\boldsymbol{Z}$, $\boldsymbol{Z}\otimes\boldsymbol{Z}^*$, and $\boldsymbol{Z}^*\otimes\boldsymbol{Z}^*$ can be derived and these results are included in Table III. In the table, $\mathrm{diag}(\cdot)$ returns the square diagonal matrix with the input column vector elements on the main diagonal [19] and zeros elsewhere.

VII. CONCLUSIONS

An introduction is given to a set of powerful tools that can be used to systematically find the derivative of complex-valued matrix functions that are dependent on complex-valued matrices. The derivation goes through the complex differential of the function and, classically, treats the differential of the complex variable and its complex conjugate as independent. This general framework is of particular interest in the many optimization problems with respect to complex parameters which arise in signal processing and communications problems.

REFERENCES

[1] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Application in Statistics and Econometrics*. Essex, England: John Wiley & Sons, Inc., 1988.

[2] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits, Syst.*, vol. CAS-25, no. 9, pp. 772–781, Sept. 1978.

[3] D. A. Harville, *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, 1997, corrected second printing, 1999.

[4] T. P. Minka. (December 28, 2000) Old and new matrix algebra useful for statistics. [Online]. Available: http://research.microsoft.com/~minka/papers/matrix/

[5] K. Kreutz-Delgado, "Real vector derivatives and gradients," Dept. of Electrical and Computer Engineering, UC San Diego, Tech. Rep. Course Lecture Supplement No. ECE275A, December 5. 2005, `http://dsp.ucsd.edu/~kreutz/PEI05.html`.

[6] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, New Jersey, USA: Prentice Hall, 2002.

[7] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, New Jersey, USA: John Wiley & Son Inc., 2003.

[8] A. van den Bos, "Complex gradient and Hessian," *Proc. IEE Vision, Image and Signal Process.*, vol. 141, no. 6, pp. 380–383, Dec. 1994.

[9] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc., Parts F and H*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[10] M. Brookes. Matrix reference manual. [Online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html

[11] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," Dept. of Electrical and Computer Engineering, UC San Diego, Tech. Rep. Course Lecture Supplement No. ECE275A, Sept.-Dec. 2005, `http://dsp.ucsd.edu/~kreutz/PEI05.html`.

[12] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[13] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[14] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[15] D. H. Jonhson and D. A. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Prentice-Hall, Inc., 1993.

[16] A. Hjørungnes and D. Gesbert. Introduction to complex-valued matrix differentiation. [Online]. Available: http://www.unik.no/personer/arehj/publications/matrixdiff.pdf

[17] ——, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, 2006, accepted for publication 11.09.2006.

[18] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press Cambridge, UK, 1991, reprinted 1999.

[19] ——, *Matrix Analysis*. Cambridge University Press Cambridge, UK, 1985, reprinted 1999.