Hessians of Scalar Functions of Complex-Valued Matrices:
A Systematic Computational Approach

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
Eurecom Institute
2229 Route des Crêtes, BP 193
F-06904 Sophia Antipolis Cédex, France
Email: gesbert@eurecom.fr

Abstract—A systematic theory is introduced for finding the four Hessians of complex-valued scalar functions with respect to a complex-valued matrix variable and the complex conjugate of this variable. It is shown how the four Hessian matrices of a scalar complex function can be identified from the second-order complex differential of the scalar function. These Hessians are the four parts of a bigger matrix which must be checked in order to identify if a stationary point is a local minimum, maximum, or saddle point. The method introduced is general such that many results can be derived using the framework introduced. Hessians are derived for some useful examples taken from signal processing related functions.

Index Terms: Hessian matrices, optimization methods, matrices.

I. INTRODUCTION

In many signal processing related problems, the unknown parameters are complex-valued matrices and, often, the task is to find the values of these complex parameters which optimize a chosen criterion. Examples of areas where these types of optimization problems might appear are telecommunications, where digital filters can contain complex-valued coefficients [1], electric circuits [2], adaptive filters [3]. For solving optimization problems, one approach is to find necessary conditions for optimality. When a scalar real-valued function depends on a complex-valued matrix parameter, the necessary conditions for optimality can be found by either setting the derivative of the function with respect to the complex-valued matrix parameter or its complex conjugate to a null vector/matrix. In [4], a theory for finding the first-order derivatives of complex-valued matrix functions which depend on complex-valued matrices was proposed. This paper provides the tools for finding Hessians, i.e., second-order derivative, in a systematic way. The proposed theory is useful when solving numerous problems which involve optimization when the unknown parameter is a complex-valued matrix. In an effort to build adaptive optimization algorithms, it is important to find out if a certain value of the complex-valued parameter matrix at a stationary point is a maximum, minimum, or saddle point and the Hessian can then be utilized very efficiently. The complex Hessian might also be used to accelerate the convergence of iterative optimization algorithms, to study the stability of iterative algorithms, and to study convexity and concavity of an objective function.

Contributions: A general framework is introduced showing how to find the four Hessians of complex-valued scalar functions with respect to the complex-valued input parameter matrix and its complex conjugate. The main contribution of this paper lies in the proposed approach for how to obtain the Hessians in a way that is both simple and systematic, based on the so-called second-order complex differential of the function. It is shown how the four Hessians must be put into a bigger matrix which must be studied for deciding the nature of a stationary point and it can also be used for deciding convexity or concavity of the objective function.

In this paper, it is assumed that the functions are twice 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 Hessian, as is classical for scalar variables [5].

Note that, the problem at hand has been treated for real-valued matrix variables in [6]. For complex-valued vectors, the Hessian matrix is briefly treated in [7], [8]. Both gradients and Hessians for scalar functions which depend on complex-valued vectors are studied in [9]. A complex version of Newton’s recursion formula is derived in [10], [11], and there the topic of Hessian matrices is briefly treated for real scalar functions which depend on complex-valued vectors. Here, a theory for finding Hessians of complex-valued matrices is proposed.

The rest of this paper is organized as follows: In Section II, the complex differential is introduced and several key differentials are presented in a table. The definition of the (first-order) derivative of complex-valued matrix functions is given in Section III. The main results of the paper is contained in Section IV, and it shows how the Hessian (second-order derivative) can be identified from the second-order differential. Some examples of how the Hessian might be calculated are also presented in Section IV. Section V contains some conclusions. The three longest proofs are put in the three appendices.

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, \((dZ)_{k,l} = d(Z)_{k,l}\). A procedure that can often be used for finding the differentials of a complex matrix function \(F : \mathbb{C}^{N \times Q} \times \mathbb{C}^{N \times Q} \rightarrow \mathbb{C}^{M \times P}\), denoted \(F(Z_0, Z_1)\), is to calculate the difference

\[
F(Z_0 + dZ_0, Z_1 + dZ_1) - F(Z_0, Z_1) = \text{First-order}(dZ_0, dZ_1) + \text{Higher-order}(dZ_0, dZ_1),
\]

where First-order(\(\cdot, \cdot\)) returns the terms that depend on either \(dZ_0\) or \(dZ_1\), of the first order, and Higher-order(\(\cdot, \cdot\)) returns the terms that depend on the higher order terms of \(dZ_0\) and \(dZ_1\). Then \(dF(Z_0, Z_1) = \text{First-order}(dZ_0, dZ_1)\), i.e., the first order term of \(F(Z_0 + dZ_0, Z_1 + dZ_1) - F(Z_0, Z_1)\). As an example, let \(F(Z_0, Z_1) = Z_0Z_1\). Then the difference in (1) is: \(F(Z_0 + dZ_0, Z_1 + dZ_1) - F(Z_0, Z_1) = Z_0(dZ_1 + (dZ_0)Z_1 + (dZ_0)dZ_1)\). Thus, \(dZ_0Z_1\) can then be identified as all the first-order terms on either \(dZ_0\) or \(dZ_1\) as \(dZ_0Z_1 = Z_0dZ_1 + (dZ_0)Z_1\).

Let \(\otimes\) and \(\circ\) denote the Kronecker and Hadamard product [12], respectively. Some of the most important basic results on complex differentials are listed in Table I, assuming \(A, B, \alpha\) to be constants, \(n \in \{1, 2, 3, \ldots\}\), and \(Z, Z_0, Z_1\) to be complex-valued matrix variables. The vectorization operator vec(\(\cdot\)) stacks the
columns of the argument matrix into a long column vector in chronological order [12]. The differentiation rule of the reshaping operator $\text{reshape()}$ is valid for any linear reshaping operator $\text{reshape()}$ of the matrix, and examples of such operators are the transpose $(\cdot)^T$ or $\text{vec}()$. 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 [6], [13] to the complex differential case.

### III. Definition of Derivative

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 the complex-valued matrix parameter $Z$ contains independent components such that these can be chosen freely.

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

$$
D_{Z} F(Z, Z^*) = \frac{d\text{vec}(F)}{dz} = \left(\partial_{z^T} \partial_{z}\right) F(z, z^*) = 
\begin{bmatrix}
\frac{\partial f_0}{\partial z_1} & \cdots & \frac{\partial f_0}{\partial z_N}, \\
\frac{\partial f_1}{\partial z_1} & \cdots & \frac{\partial f_1}{\partial z_N}, \\
\vdots & \ddots & \vdots \\
\frac{\partial f_M-1}{\partial z_1} & \cdots & \frac{\partial f_M-1}{\partial z_N}
\end{bmatrix}
\Delta_z.
$$

(3)

where $z_1$ and $f_i$ is component number $i$ of the vectors $z$ and $f$, respectively.

By studying the way all the derivatives are arranged in Definition 2 and Definition 1, it can be shown that the connection between these two definitions is given by the following two relations:

$$
D_{Z} F(Z, Z^*) = \frac{\partial \text{vec}(F(Z, Z^*))}{\partial z} = \left(\partial_{z^T} \partial_{z}\right) F(z, z^*),
$$

(4)

$$
D_{Z^*} F(Z, Z^*) = \frac{\partial \text{vec}(F(Z, Z^*))}{\partial z^*} = \left(\partial_{z_T^*} \partial_{z^*}\right) F(z, z^*).\tag{5}
$$

(4) and (5) show how the all $MPNQ$ partial derivatives of all the respective components of $F$ with respect to all the components of $Z$ and $Z^*$ are arranged in $D_{Z} F(Z, Z^*)$ and $D_{Z^*} F(Z, Z^*)$ when using the notation introduced in Definition 2.

### IV. Hessians of Complex-Valued Scalar Functions

The Hessian matrices of a function are matrices containing the second-order derivatives of the function. In this section, the Hessian matrices are defined and it is shown how they can be obtained. Only the case of scalar functions $f \in \mathbb{C}$ is treated, since this is the case of interest in most practical situations. However, the results can be extended to the vector- and matrix-functions as well. The way the Hessian is defined here follows in the same lines as given in [6], treating the real case.

#### A. Useful Results

In this section, some useful results will be given which are used for identifying the Hessians.

**Lemma 1:** Let $A, B \in \mathbb{C}^{N \times N}, z^T A z = z^T B z \forall z \in \mathbb{C}^{N \times 1}$ is equivalent to $A + A^T = B + B^T$.

The proof of Lemma 1 can be found in Appendix I.

**Corollary 1:** Let $A \in \mathbb{C}^{N \times N}, z^T A z = 0 \forall z \in \mathbb{C}^{N \times 1}$ is equivalent to $A^T = -A$.

**Proof:** Set $B = 0_{N \times N}$ in Lemma 1, and the result follows.

**Lemma 2:** Let $A, B \in \mathbb{C}^{N \times N}, Z^* A Z = 0 \forall z \in \mathbb{C}^{N \times 1}$ is equivalent to $A = B$.

The proof of Lemma 2 can be found in Appendix II.

**Lemma 3:** Let $Z \in \mathbb{C}^{N \times Q}$ and let $B_i \in \mathbb{C}^{NQ \times NQ}$. If $(d\text{vec}(Z)) B_0 d\text{vec}(Z) = (d\text{vec}(Z)) B_1 d\text{vec}(Z) = (d\text{vec}(Z)) B_2 d\text{vec}(Z) = 0$ for all $dZ \in \mathbb{C}^{N \times Q}$, then $B_0 = -B_0^T, B_1 = 0_{NQ \times NQ}$, and $B_2 = -B_2^T$.

The proof of Lemma 3 can be found in Appendix III.

#### B. Identification of the Complex Hessian Matrices

The second-order differential can be calculated in order to identify the Hessian matrices. Neither of the matrices $dZ$ nor $dZ^*$ is a function of $Z$ or $Z^*$ and, hence, their differentials are the zero matrix. Mathematically, this can be formulated as:

$$
d^2 Z = (dZ) = 0_{N \times Q} = (dZ^*) = 0_{N \times Q}.
$$

If $f \in \mathbb{C}$, then $(d^2 f)^T = (df)^T = d^2 f = d^2 f$ and if $f \in \mathbb{R}$, then $(d^2 f)^H = (df)^H = d^2 f$.

The Hessian depends on two variables such that the notation must include which variable the Hessian is calculated with respect to. If the Hessian is calculated with respect to $Z_0$ and $Z_1$, the Hessian is denoted $H_{Z_0, Z_1} f$. The following definition is used for $H_{Z_0, Z_1} f$ and it is an extension of the definition in [6, page 189], to complex scalar functions.

**Definition 3:** Let $Z_i \in \mathbb{C}^{N_i \times Q_i}, i \in \{0, 1\}$, then $H_{Z_0, Z_1} f \in \mathbb{C}^{N_0 \times Q_1 \times N_1 \times Q_0}$ and $H_{Z_0, Z_1} f \Delta_{Z_0} (D_{Z_1} f)^T$.

Later, the following proposition is important for showing the symmetry of the complex Hessian matrix.

**Proposition 1:** Let $f : \mathbb{C}^{N \times N} \times \mathbb{C}^{N \times Q} \rightarrow \mathbb{C}$. Assumed that the function $f(Z, Z^*)$ is twice differentiable with respect to all of the variables inside $Z$ and $Z^*$, when these variables are treated as independent variables. Then $[6] \frac{\partial^2 f}{\partial z_i \partial z_j} = \frac{\partial^2 f}{\partial z_i \partial z_j},$ and $\frac{\partial^2 f}{\partial z_i \partial z_j} = \frac{\partial^2 f}{\partial z_i \partial z_j},$ where $m, k \in \{0, 1, \ldots, N-1\}$ and $n, l \in \{0, 1, \ldots, Q-1\}$.

Let $p_i = N_0 k_i + l_i$ where $i \in \{0, 1\}, k_i \in \{0, 1, \ldots, Q_i - 1\}$ and $l_i \in \{0, 1, \ldots, N_i - 1\}$. As a consequence of Definition 3 and (4), it follows that element number $(p_0, p_1)$ of $H_{Z_0, Z_1} f$ is given by:
As an immediate consequence of (7) and Proposition 1, it follows that for twice differentiable functions:

\[(H_z, z f)^T = H_z, z f, (H_{z z}, z^2 f)^T = H_{z z}, z^2 f, (H_{z z^2}, z^2 f)^T = H_{z z}, z f.\]

In order to find an identification equation for the Hessians of the function \( f \) with respect to all the possible combinations of the variables \( Z \) and \( Z^* \), an appropriate form of the expression \( d^2 f \) is required. This expression is now derived. From (2), the first-order differential of the function \( f \) can be written as:

\[df = (D_z f)vec(Z) + (D_{z^2} f)vec(Z^*),\]

where \( D_z f \) and \( D_{z^2} f \) are the complex Hessian matrices of the scalar function \( f \). Manipulate the differential operator twice to the latest expression of \( f \) where \( \frac{\partial Z}{\partial Z}\in C_{1 \times NQ}^+ \) and \( \frac{\partial Z}{\partial Z} \) again can be expressed by the use of (2) as:

\[(dD_z f)^T = [D_z (D_z f)^T] dvec(Z) + [D_{z z} (D_z f)^T] dvec(Z^*),\]

\[dD_{z^2} f = [dvec^T(Z)] [D_z (D_{z^2} f)^T]^T dvec(Z) + [dvec^T(Z^*)] [D_{z z} (D_z f)^T] dvec(Z^*).\]

\[d^2 f = (dD_z f)^T dvec(Z) + (dD_{z^2} f) dvec(Z^*)\]

\[= [dvec^T(Z)] [D_z (D_z f)^T]^T dvec(Z) + [dvec^T(Z^*)] [D_z (D_{z^2} f)^T] dvec(Z) + [dvec^T(Z)] [D_{z z} (D_{z^2} f)^T] dvec(Z^*) + [dvec^T(Z^*)] [D_{z z} (D_z f)^T] dvec(Z)\]

\[= [dvec^T(Z), dvec^T(Z^*)] [D_z (D_z f)^T, D_z (D_{z^2} f)^T, D_{z z} (D_{z^2} f)^T, D_{z z} (D_z f)^T]^T dvec(Z)\]

From Definition 3, it follows that \( d^2 f \) can be rewritten as:

\[d^2 f = [dvec^T(Z), dvec^T(Z^*)] \begin{bmatrix} \mathcal{H}_{z z}, z f, \mathcal{H}_{z z}, z^2 f \\ \mathcal{H}_{z z}, z f, \mathcal{H}_{z z}, z^2 f \end{bmatrix} [dvec(Z), dvec(Z^*)].\]
importance of checking the whole matrix \[ \begin{bmatrix} \mathcal{H}_z z \cdot z \cdot f \, & \mathcal{H}_z z \cdot z \cdot f \end{bmatrix} \]
when deciding whether or not a stationary point is a local minimum, local maximum, or a saddle point. Figure 1 shows \( f \) for \( N = Q = 1 \), and it is seen that the origin is indeed a saddle point.

V. Conclusions

A way of identifying the Hessian of complex-valued scalar functions which depend on complex-valued matrices has been proposed. Some examples were included showing how to find the Hessian of signal processing related problems. It is shown that in order to decide if a stationary point is a local maximum, minimum, or saddle point, a bigger matrix containing the four Hessians must be studied, and the same matrix can be used to decide convexity or concavity.

APPENDIX I

Proof of Lemma 1

Proof: Let \((A)_{k,l} = a_{k,l}\) and \((B)_{k,l} = b_{k,l}\). Assume first that \(z^T A z = z^T B z \) \( \forall \) \( z \in \mathbb{C}^{N \times 1} \), and set \( z = e_k \) where \( k \in \{0, 1, \ldots, N - 1\} \). Then \( e_k^T A e_k = e_k^T B e_k \), giving that \( a_{k,k} = b_{k,k} \) for all \( k \in \{0, 1, \ldots, N - 1\} \). Setting \( z = e_k + e_l \) leads to \((e_k^T + e_l^T) A(e_k + e_l) = (e_k^T + e_l^T) B(e_k + e_l)\), which results in \( a_{k,l} + a_{l,k} = b_{k,l} + b_{l,k} \). Eliminating equal terms gives \( a_{k,l} + a_{l,k} = b_{k,l} + b_{l,k} \) which can be written \( A + A^T = B + B^T \).

Assuming \( A + A^T = B + B^T \), leads to \( z^T A z = \frac{1}{2} (z^T A z + z^T A^T z) = \frac{1}{2} z^T (A + A^T) z = \frac{1}{2} z^T (B + B^T) z = \frac{1}{2} (z^T B z + z^T B^T z) = z^T B z \) for all \( z \in \mathbb{C}^{N \times 1} \).

APPENDIX II

Proof of Lemma 2

Proof: Let \((A)_{k,l} = a_{k,l}\) and \((B)_{k,l} = b_{k,l}\). Assume first that \(z^H A z = z^H B z \) \( \forall \) \( z \in \mathbb{C}^{N \times 1} \), and set \( z = e_k \) where \( k \in \{0, 1, \ldots, N - 1\} \). This gives that \( a_{k,k} = b_{k,k} \) for all \( k \in \{0, 1, \ldots, N - 1\} \). Also in the same way as in the proof of Lemma 1 setting \( z = e_k + e_l \), then manipulations give \( A + A^T = B + B^T \). The equations \( A + A^T = B + B^T \) and \( A - A^T = B - B^T \) imply that \( A = B \).

If \( A = B \), it follows that \( z^H A z = z^H B z \) for all \( z \in \mathbb{C}^{N \times 1} \).

APPENDIX III

Proof of Lemma 3

Proof: Substitution of \( d \text{vec}(Z) = d \text{vec} \{\text{Re} \{Z\} \} + j d \text{vec}(\text{Im} \{Z\}) \) and \( d \text{vec}(Z^*) = d \text{vec}(\text{Re} \{Z\}) - j d \text{vec}(\text{Im} \{Z\}) \) into the second-order differential expression given in the lemma leads to:

\[
\begin{align*}
&\left\{ d \text{vec}(\text{Re} \{Z\}) \right\} \left[ B_0 + B_1 + B_2 \right] d \text{vec}(\text{Re} \{Z\}) \\
&\left\{ d \text{vec}(\text{Im} \{Z\}) \right\} \left[ -B_0 + B_1 - B_2 \right] d \text{vec}(\text{Im} \{Z\}) \\
&\left\{ d \text{vec}(\text{Re} \{Z\}) \right\} \left[ j \left( B_0 + B_1 + B_2 \right) + j \left( B_1 - B_2 \right) \right] \\
&\left\{ d \text{vec}(\text{Im} \{Z\}) \right\} \left[ -j \left( B_2 + B_1 \right) \right] d \text{vec}(\text{Im} \{Z\}) = 0.
\end{align*}
\]

(17) is valid for all \( dZ \) and the differentials of \( d \text{vec}(\text{Re} \{Z\}) \) and \( d \text{vec}(\text{Im} \{Z\}) \) are independent. If \( d \text{vec}(\text{Im} \{Z\}) \) is set to the zero vector, then it follows from (17) and Corollary 1 that:

\[
B_0 + B_1 + B_2 = -B_0^T - B_1^T - B_2^T.
\]

(18) In the same way, setting \( d \text{vec}(\text{Re} \{Z\}) \) to the zero vector, then it follows from (17) and Corollary 1 that:

\[
-B_0 + B_1 - B_2 = B_0^T - B_1^T + B_2^T.
\]

(19)

Because of the skew symmetry in (18) and (19) and the linear independence of \( d \text{vec}(\text{Re} \{Z\}) \) and \( d \text{vec}(\text{Im} \{Z\}) \), it follows from (17) and Corollary 1 that:

\[
\left( B_0 + B_0^T \right) + \left( B_1 - B_1^T \right) - \left( B_2 + B_2^T \right) = 0_{NQ \times NQ}.
\]

(20)

APPENDIX IV

Proof of Lemma 4

Proof: Let \( (A)_{k,l} = a_{k,l} \) and \( (B)_{k,l} = b_{k,l} \). Assume first that \( z^T A z = z^T B z \) \( \forall \) \( z \in \mathbb{C}^{N \times 1} \), and set \( z = e_k \) where \( k \in \{0, 1, \ldots, N - 1\} \). This gives that \( a_{k,k} = b_{k,k} \) for all \( k \in \{0, 1, \ldots, N - 1\} \). Also in the same way as in the proof of Lemma 1 setting \( z = e_k + e_l \), then manipulations give \( A + A^T = B + B^T \). The equations \( A + A^T = B + B^T \) and \( A - A^T = B - B^T \) imply that \( A = B \).

If \( A = B \), it follows that \( z^H A z = z^H B z \) for all \( z \in \mathbb{C}^{N \times 1} \).

APPENDIX V

Proof of Lemma 5

Proof: Substitution of \( d \text{vec}(Z) = d \text{vec}(\text{Re} \{Z\}) + j d \text{vec}(\text{Im} \{Z\}) \) into the second-order differential expression gives the lemma leads to:

\[
\begin{align*}
&\left\{ d \text{vec}(\text{Re} \{Z\}) \right\} \left[ B_0 + B_1 + B_2 \right] d \text{vec}(\text{Re} \{Z\}) \\
&\left\{ d \text{vec}(\text{Im} \{Z\}) \right\} \left[ -B_0 + B_1 - B_2 \right] d \text{vec}(\text{Im} \{Z\}) \\
&\left\{ d \text{vec}(\text{Re} \{Z\}) \right\} \left[ j \left( B_0 + B_1 + B_2 \right) + j \left( B_1 - B_2 \right) \right] \\
&\left\{ d \text{vec}(\text{Im} \{Z\}) \right\} \left[ -j \left( B_2 + B_1 \right) \right] d \text{vec}(\text{Im} \{Z\}) = 0.
\end{align*}
\]

(17) is valid for all \( dZ \) and the differentials of \( d \text{vec}(\text{Re} \{Z\}) \) and \( d \text{vec}(\text{Im} \{Z\}) \) are independent. If \( d \text{vec}(\text{Im} \{Z\}) \) is set to the zero vector, then it follows from (17) and Corollary 1 that:

\[
B_0 + B_1 + B_2 = -B_0^T - B_1^T - B_2^T.
\]

(18) In the same way, setting \( d \text{vec}(\text{Re} \{Z\}) \) to the zero vector, then it follows from (17) and Corollary 1 that:

\[
-B_0 + B_1 - B_2 = B_0^T - B_1^T + B_2^T.
\]

(19)

Because of the skew symmetry in (18) and (19) and the linear independence of \( d \text{vec}(\text{Re} \{Z\}) \) and \( d \text{vec}(\text{Im} \{Z\}) \), it follows from (17) and Corollary 1 that:

\[
\left( B_0 + B_0^T \right) + \left( B_1 - B_1^T \right) - \left( B_2 + B_2^T \right) = 0_{NQ \times NQ}.
\]

(20)