Article

Extensions of Some Statistical Concepts to the Complex Domain

Arak M. Mathai [ID]

Mathematics and Statistics, McGill University, Montreal, QC H3A 2K6, Canada; directorcms458@gmail.com

Abstract: This paper deals with the extension of principal component analysis, canonical correlation analysis, the Cramer–Rao inequality, and a few other statistical concepts in the real domain to the corresponding complex domain. Optimizations of Hermitian forms under a linear constraint, a bilinear form under Hermitian-form constraints, and similar maxima/minima problems in the complex domain are discussed. Some vector/matrix differential operators are developed to handle the above types of problems. These operators in the complex domain and the optimization problems in the complex domain are believed to be new and novel. These operators will also be useful in maximum likelihood estimation problems, which will be illustrated in the concluding remarks. Detailed steps are given in the derivations so that the methods are easily accessible to everyone.

Keywords: principal component analysis; canonical correlation analysis; extensions to complex domain; linear forms; Hermitian forms; vector; matrix differential operators; Cramer–Rao inequality in the complex domain

MSC: 62H25; 62H20; 62J10; 15B52; 15B57

1. Introduction

In most textbooks on statistics, scalar/vector/matrix-variate random variables in the complex domain and the corresponding statistical analysis in the complex domain, are not discussed. But, in a large number of physical situations, it is natural or more convenient to represent variables in the complex domain. Hence, statistical techniques in the complex domain are required for data analysis in these situations. In the physical science and engineering literature, there are a number of papers dealing with random variables in the complex domain. Data reduction techniques such as principal component analysis in the complex domain seem to be the main topic in these areas. Most of the papers in these applied areas, concentrate on developing algorithms for computing eigenvalues and eigenvectors, which are useful and relevant in principal component analysis, independent component analysis, factor analysis, and so on. Statistical analysis in the complex domain is widely used in the analysis of multi-look return signals in radar [1], in multi-task learning in artificial intelligence and machine learning [2], in problems such as signal processing [3], in principal component analysis and independent component analysis in analyzing meteorological data in the complex domain [4], in optimal allocation of resources, especially energy resources [5], in holography, microscopy and optical metrology [6], in delayed mixing in speech processing, in biomedical signal analysis, and in financial data modeling, etc. [7].

In the present paper, vector/matrix differential operators in the complex domain are defined. Then, these are applied in optimizing a Hermitian form under a linear constraint, a bilinear form under Hermitian-form constraints, etc. Then, as applications of these optimization problems, the real-domain techniques of principal component analysis and canonical correlation analysis are extended to the complex domain. Other statistical concepts such as the Cramer–Rao inequality, least square procedure, and related aspects are extended to the complex domain. Detailed derivations are given so that the methods will be accessible even to beginners.

The following notation is used in this paper: Scalar variables, whether mathematical or random, are denoted by lower-case letters such as x, y. Vector/matrix variables are denoted by capital letters such as X, Y. Scalar constants are denoted by a, b, etc., and vector, matrix constants by A, B, etc. The wedge product of the differentials $\mathrm{d}x$ and $\mathrm{d}y$ is defined as $\mathrm{d}x \wedge \mathrm{d}y = -\mathrm{d}y \wedge \mathrm{d}x$, where x and y are two real scalar variables, so that $\mathrm{d}x \wedge \mathrm{d}x = 0, \mathrm{d}y \wedge \mathrm{d}y = 0$. Let $X = (x_{jk})$ be an $m \times n$ matrix with distinct real scalar variables x_{jk} as elements, then $\mathrm{d}X = \wedge_{j=1}^{m} \wedge_{k=1}^{n} \mathrm{d}x_{jk}$. The transpose of a matrix X is denoted by a prime as X'. For a $p \times p$ matrix X, if $X = X'$ (symmetric), then $\mathrm{d}X = \wedge_{j \geq k} \mathrm{d}x_{jk} = \wedge_{j \leq k} \mathrm{d}x_{jk}$. Variables in the complex domain are written with a tilde, such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$. If $\tilde{X}$ is a $p \times q$ matrix in the complex domain, then $\tilde{X} = X_1 + iX_2, i = \sqrt{(-1)}, X_1, X_2$ are real, then $\mathrm{d}\tilde{X}$ is defined as $\mathrm{d}\tilde{X} = \mathrm{d}X_1 \wedge \mathrm{d}X_2$. The determinant of a real $p \times p$ matrix Y is written as $|Y|$ or as $\det(Y)$ and if $\tilde{Y}$ is in the complex domain, then the absolute value of the determinant of $\tilde{Y}$ is written as $|\det(Y)|$. Also, $\mathrm{tr}(Y)$ means the trace of the square matrix Y. A $p \times p$ real matrix A being positive definite is written as $A > O$, and $\tilde{X} = \tilde{X}^* > O$ indicates that the matrix $\tilde{X}$, in the complex domain, is Hermitian positive definite. Other notation is explained when it occurs for the first time.

This paper is organized as follows: Section 2 starts with defining a vector differential operator in the complex domain. Then, with the help of this operator, some optimization problems such as optimizing a linear form subject to Hermitian-form constraint, a Hermitian form under Hermitian-form constraint, and a bilinear form with Hermitian-form constraints are discussed. Then, a matrix differential operator in the complex domain is defined. Differentiations of the trace of a product of matrices and determinant of a matrix, by using the matrix differential operator in the complex domain, are dealt with. Section 3 deals with the extension of principal component analysis to the complex domain. Section 4 delves into the extension of canonical correlation analysis to the complex domain. Section 5 examines the extension of the Cramer–Rao inequality, Cauchy–Schwarz inequality, and the least square estimation procedure to the complex domain. Detailed steps are given in each situation.

2. Optimization Involving Linear Forms, Traces, and Determinants

Here, we will consider linear forms and their differentiations first, and then, we will consider some situations of optimizing a Hermitian form with linear constraint and optimization of a linear form with Hermitian-form constraint. We consider some basic results to start with. Let

$$
A = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_p \end{bmatrix}, A = A_1 + iA_2, i = \sqrt{(-1)}, \tilde{X} = X_1 + iX_2
$$

where A_1, A_2, X_1, X_2 are real vectors. Assume that A is a non-null arbitrary coefficient vector and $\tilde{X}$ is a known vector random variable in the complex domain.

Theorem 1. *For A and $\tilde{X}$ as defined above,*

$$
\tilde{u} = A^* \tilde{X} \Rightarrow \frac{\partial \tilde{u}}{\partial A} = 2\tilde{X}. \tag{1}
$$

Proof. This can be easily seen from the following: $\tilde{u} = (A_1' - iA_2')(X_1 + iX_2) = A_1'X_1 + A_2'X_2 + i(A_1'X_2 - A_2'X_1)$ Then,

$$
\frac{\partial \tilde{u}}{\partial A_1} = X_1 + iX_2 \text{ and } \frac{\partial \tilde{u}}{\partial A_2} = X_2 - iX_1 \tag{2}
$$

from the corresponding real variable case. That is,

$$\left(\frac{\partial}{\partial A_1} + i\frac{\partial}{\partial A_2}\right)\tilde{u} = \frac{\partial \tilde{u}}{\partial A} = 2(X_1 + iX_2) = 2\tilde{X}.$$

□

Now, we consider a slightly more general result. Consider $\tilde{u} = A^*\Sigma_{11}\tilde{X}, \Sigma_{11} = \Sigma_{11}^* > O$. Let $\frac{\partial}{\partial A^c} = \left(\frac{\partial}{\partial A_1} - i\frac{\partial}{\partial A_2}\right)$ and $\frac{\partial}{\partial A^*} = \left(\frac{\partial}{\partial A_1'} - i\frac{\partial}{\partial A_2'}\right)$. Then, from (2) note that $\frac{\partial \tilde{u}}{\partial A^c} = O$ and $\frac{\partial \tilde{u}}{\partial A^*} = O$; $\frac{\partial \tilde{u}}{\partial \tilde{X}} = O$; $\frac{\partial \tilde{u}}{\partial \tilde{X}^c} = 2A$ and $\frac{\partial \tilde{u}}{\partial \tilde{X}^*} = 2A'$, where A^c means the complex conjugate of A.

Theorem 2. *For $\tilde{X}$ and A as defined above, let $\tilde{u} = A^*\Sigma_{11}\tilde{X}$. Let Σ_{11} be a $p \times p$ matrix, free of the elements in A and $\tilde{X}$. Then,*

$$\frac{\partial \tilde{u}}{\partial A} = 2\Sigma_{11}\tilde{X}. \tag{3}$$

Proof. Since Σ_{11} and $\tilde{X}$ are free of the elements in A, we may take $\Sigma_{11}\tilde{X} = B = B_1 + iB_2$, $i = \sqrt{(-1)}, B_1, B_2$ real. Then, from Theorem 1 the result follows. □

Now, consider a $p \times 1$ vector A and a $p \times p$ matrix Σ and the Hermitian form $\tilde{u} = A^*\Sigma A$ in the complex domain where $\Sigma = \Sigma^* > O$ is free of the elements in A. Then, we have the following result:

Theorem 3. *For the Hermitian form as defined above, where $A = A_1 + iA_2, \Sigma = \Sigma_1 + i\Sigma_2$, $\Sigma = \Sigma^* > O, A_1, A_2, \Sigma_1, \Sigma_2$ are real,*

$$\frac{\partial}{\partial A}[\tilde{u}] = \frac{\partial}{\partial A}[A^*\Sigma A] = 2\Sigma A. \tag{4}$$

Proof. Opening up $\tilde{u}$, we have the following:

$$\tilde{u} = A^*\Sigma A = (A_1' - iA_2')\Sigma(A_1 + iA_2) = A_1'\Sigma A_1 + A_2'\Sigma A_2 + i(A_1'\Sigma A_2 - A_2'\Sigma A_1). \tag{5}$$

Since $\Sigma = \Sigma^*$, we have $\Sigma_1 = \Sigma_1'$ (real symmetric), $\Sigma_2 = -\Sigma_2'$ (real skew symmetric). $A_1'\Sigma A_1 = A_1'\Sigma_1 A_1 + iA_1'\Sigma_2 A_1 = A_1'\Sigma_1 A_1$ because $A_1'\Sigma_2 A_1 = 0$ due to Σ_2 being real skew symmetric. Similarly, $A_2'\Sigma A_2 = A_2'\Sigma_1 A_2$. Then, from the results in the real case, we have the following:

$$\frac{\partial}{\partial A_1}[A_1'\Sigma_1 A_1] = 2\Sigma_1 A_1, \ \frac{\partial}{\partial A_2}[A_2'\Sigma_1 A_2] = 2\Sigma_1 A_2. \tag{6}$$

Consider

$$i(A_1'\Sigma A_2 - A_2'\Sigma A_1) = i[(A_1'\Sigma_1 A_2 - A_2'\Sigma_1 A_1) + i(A_1'\Sigma_2 A_2 - A_2'\Sigma_2 A_1)].$$

But, $\Sigma_1 = \Sigma_1'$ and $(A_2'\Sigma_1 A_1)' = A_1'\Sigma_1 A_2$ and real 1×1, and hence, $A_1'\Sigma_1 A_2 - A_2'\Sigma_1 A_1 = 0$. Also, $A_2'\Sigma_2 A_1 = -A_1'\Sigma_2 A_2$ because $\Sigma_2 = -\Sigma_2'$, and hence, $A_1'\Sigma_2 A_2 - A_2'\Sigma_2 A_1 = 2A_1'\Sigma_2 A_2 = -2A_2'\Sigma_2 A_1$. Therefore,

$$i(A_1'\Sigma A_2 - A_2'\Sigma A_1) = i^2 2A_1'\Sigma_2 A_2 = -i^2 2A_2'\Sigma_2 A_1.$$

Then, from the real case,

$$\frac{\partial}{\partial A_1}[i^2 2 A_1' \Sigma_2 A_2] = 2i^2 \Sigma_2 A_2 = 2(i\Sigma_2)(iA_2) \Rightarrow$$

$$\frac{\partial}{\partial A_1}(A^*\Sigma A) = 2\Sigma_1 A_1 + 2(i\Sigma_2)(iA_2)$$

$$\frac{\partial}{\partial A_2}(-2i^2 A_2' \Sigma_2 A_1) = -2i^2 \Sigma_2 A_1 = 2\Sigma_2 A_1 \Rightarrow$$

$$\frac{\partial}{\partial A_2}(A^*\Sigma A) = 2\Sigma_1 A_2 + 2\Sigma_2 A_1 \tag{7}$$

from the corresponding real case. Then, from (2)–(5),

$$\begin{aligned}
\frac{\partial}{\partial A}[A^*\Sigma A] &= (\frac{\partial}{\partial A_1} + i\frac{\partial}{\partial A_2})[A^*\Sigma A] \\
&= [2\Sigma_1 A_1 + 2(i\Sigma_2)(iA_2)] + i[2\Sigma_1 A_2 + 2\Sigma_2 A_1] \\
&= 2\Sigma_1 A_1 + i2\Sigma_1 A_2 + 2i\Sigma_2(A_1 + iA_2) = 2\Sigma_1 A + 2i\Sigma_2 A = 2\Sigma A.
\end{aligned}$$

This establishes the result. □

2.1. Optimization of a Linear Form Subject to Hermitian-Form Constraint

Consider a linear form $\tilde{u} = A^*\tilde{X}$, where A and $\tilde{X}$ are $p \times 1$ non-null vectors, with A being arbitrary, and $\tilde{X}$ being a known vector variable with variance $\text{Var}(\tilde{u}) = A^*\Sigma_{11} A$, where $\Sigma_{11} = \Sigma_{11}^* > O$ is the covariance matrix in $\tilde{X}$. Consider the optimization of $\tilde{u}$, subject to the constraint that the variance of $\tilde{u}$ is fixed, say unity, that is, $A^*\Sigma_{11} A = 1$. Then, we have the following result:

Theorem 4. *For the linear form $\tilde{u} = A^*\tilde{X}$ and the constraint $A^*\Sigma_{11} A = 1$, as defined above,*

$$\max_{A^*\Sigma_{11}A=1}[A^*\tilde{X}] = \sqrt{\tilde{X}^*\Sigma_{11}^{-1}\tilde{X}}. \tag{8}$$

Proof. Let $\tilde{w} = A^*\tilde{X} - \lambda(A^*\Sigma_{11} A - 1)$, where λ is a Lagrangian multiplier. Observe that $\Sigma_{11} = \Sigma_{11}^*$ and we assume that it is Hermitian positive definite. From previous results (1) and (2),

$$\frac{\partial \tilde{w}}{\partial A} = 2\tilde{X} - \lambda 2\Sigma_{11} A = O \Rightarrow \tilde{X} = \lambda \Sigma_{11} A \Rightarrow \lambda = A^*\tilde{X}. \tag{9}$$

That is,

$$A = \frac{1}{\lambda}\Sigma_{11}^{-1}\tilde{X} \Rightarrow A^* = \frac{1}{\lambda^c}\tilde{X}^*\Sigma_{11}^{-1} \Rightarrow \lambda = A^*\tilde{X} = \frac{1}{\lambda^c}\tilde{X}^*\Sigma_{11}^{-1}\tilde{X}. \tag{10}$$

But from (9), $\lambda = A^*\tilde{X}$ which means that the maximum of our linear form is the largest λ and the minimum of our linear form is the smallest λ. But, from (9) and (10), $\lambda\lambda^c = |\lambda|^2 = \tilde{X}^*\Sigma_{11}^{-1}\tilde{X}$, where $|\lambda|$ means the absolute value of λ and λ^c is the complex conjugate of λ.

$$\max_{A^*\Sigma_{11}A=1}[\tilde{u}] = \sqrt{\tilde{X}^*\Sigma_{11}^{-1}\tilde{X}}, \quad \tilde{X}^*\Sigma_{11}^{-1}\tilde{X} > 0$$

being positive definite Hermitian form. This completes the proof. □

2.2. Optimization of a Hermitian Form Subject to Linear Constraint

Now, consider a problem of optimizing $A^*\Sigma_{11} A$ subject to $A^*\tilde{X} = \alpha$ fixed, where A and $\tilde{X}$ are $p \times 1$ vectors and $\Sigma_{11} = \Sigma_{11}^* > O$ is a $p \times p$ Hermitian positive definite matrix, free of the elements of A. Then, we have the following result:

Theorem 5. *For the Hermitian form $A^*\Sigma_{11}A$ and linear form constraint $A^*\tilde{X} = \alpha$, we have*

$$\max_{A^*\tilde{X}=\alpha} [A^*\Sigma_{11}A] = +\infty, \ \min_{A^*\tilde{X}=\alpha} [A^*\Sigma_{11}A] = \frac{|\alpha|^2}{\tilde{X}^*\Sigma_{11}^{-1}\tilde{X}}. \tag{11}$$

Proof. Let λ be a Lagrangian multiplier and consider $\tilde{w} = A^*\Sigma_{11}A - \lambda(A^*\tilde{X} - \alpha)$. From (1) and (2), we have

$$\frac{\partial \tilde{w}}{\partial A} = 2\Sigma_{11}A - 2\lambda\tilde{X} \Rightarrow \Sigma_{11}A = \lambda\tilde{X} \Rightarrow A^*\Sigma_{11}A = \lambda A^*\tilde{X} = \lambda\alpha. \tag{12}$$

From (12),

$$A = \lambda\Sigma_{11}^{-1}\tilde{X} \Rightarrow \alpha = A^*\tilde{X} = \lambda^c \tilde{X}^*\Sigma_{11}^{-1}\tilde{X} \Rightarrow |\lambda| = \frac{|\alpha|}{\tilde{X}^*\Sigma_{11}^{-1}\tilde{X}} \tag{13}$$

because $\tilde{X}^*\Sigma_{11}^{-1}\tilde{X}$ is Hermitian positive definite, where $|\lambda|$ and $|\alpha|$ denote the absolute values of λ and α, respectively. Then, $\max[\tilde{X}^*A\tilde{X}] = +\infty$ since A is arbitrary and since the linear restriction cannot eliminate the effect of A fully. Hence, we look for the minimum. From (13),

$$|\lambda| = \frac{|\alpha|}{\tilde{X}^*\Sigma_{11}^{-1}\tilde{X}} \Rightarrow |\lambda\alpha| = \frac{|\alpha|^2}{\tilde{X}^*\Sigma_{11}^{-1}\tilde{X}}.$$

Hence,

$$\min_{A^*\tilde{X}=1} [A^*\Sigma_{11}A] = \frac{|\alpha|^2}{\tilde{X}^*\Sigma_{11}^{-1}\tilde{X}}.$$

This completes the proof. $\square$

2.3. Differentiation of Traces of Matrix Products and Matrix Differential Operator

Now, we consider a few matrix-variate cases, where also the problem is essentially optimization involving vector variable situations. Let $A = (a_{jk})$ be a $p \times q$ matrix of arbitrary elements a_{jk}. Let $\tilde{X} = (\tilde{x}_{jk})$ be a $p \times q$ matrix in the complex domain with distinct scalar complex variables $\tilde{x}_{jk}$ as elements. Consider $\tilde{u} = \text{tr}(A^*\tilde{X})$. Let A_j and $\tilde{X}_j$ be the j-th columns of A and $\tilde{X}$, respectively. Then, $\tilde{u} = \sum_{j=1}^q A_j^*\tilde{X}_j$. Let $A_j = A_{j1} + iA_{j2}$, $i = \sqrt{(-1)}$, A_{j1}, A_{j2} are real $p \times 1$ vectors. Let $\tilde{u}_j = A_j^*\tilde{X}_j$. Consider the operator

$$\frac{\partial}{\partial A_j} = (\frac{\partial}{\partial A_{j1}} + i\frac{\partial}{\partial A_{j2}}), \ \frac{\partial}{\partial A_j^*} = (\frac{\partial}{\partial A_{j1}'} - i\frac{\partial}{\partial A_{j2}'}), \ \frac{\partial}{\partial A_j^c} = (\frac{\partial}{\partial A_{j1}} - i\frac{\partial}{\partial A_{j2}})$$

and then, $\frac{\partial \tilde{u}_j}{\partial A_j} = \frac{\partial \tilde{u}}{\partial A_j}$ since we are taking partial derivatives. Then, we have the following result:

Theorem 6. *Let $\tilde{u}, A, A_j, \tilde{X}_j$ be as defined above. Then,*

$$\frac{\partial \tilde{u}}{\partial A} = \frac{\partial}{\partial A}[\text{tr}(A^*\tilde{X})] = 2\tilde{X}. \tag{14}$$

Proof. From the previous result,

$$\frac{\partial \tilde{u}_j}{\partial A_j} = (\frac{\partial}{\partial A_{j1}} + i\frac{\partial}{\partial A_{j2}})\tilde{u}_j$$

$$= (\frac{\partial}{\partial A_{j1}} + i\frac{\partial}{\partial A_{j2}})(A_j^*\tilde{X}_j) = 2\tilde{X}_j, j = 1, ..., q,$$

from Theorem 1. But, $\frac{\partial \tilde{u}_j}{\partial A_j} = \frac{\partial \tilde{u}}{\partial A_j}$. Now, stack up these side by side as q columns for $j = 1, ..., q$. Note that $A = [A_1, ..., A_q]$, $\frac{\partial}{\partial A} = [\frac{\partial}{\partial A_1}, ..., \frac{\partial}{\partial A_q}]$, and $\frac{\partial \tilde{u}}{\partial A_j} = 2\tilde{X}_j$, and hence, $\frac{\partial \tilde{u}}{\partial A} = 2\tilde{X}$. This completes the proof. □

For the next result to be established, we need a result on matrices, which will be stated here as a lemma.

Lemma 1. *For two $p \times p$ real matrices A and B, where $A = A'$ (symmetric) and $B = -B'$ (skew symmetric), $\mathrm{tr}(A'B) = 0$.*

Proof. For two $p \times p$ matrices A_1 and B_1, it is well known that $\mathrm{tr}(A_1) = \mathrm{tr}(A_1')$ and $\mathrm{tr}(A_1 B_1) = \mathrm{tr}(B_1 A_1)$. Then, for our matrices A and B, $A = A', B = -B'$, we have $\mathrm{tr}(A'B) = \mathrm{tr}(A'B)' = \mathrm{tr}(B'A) = -\mathrm{tr}(BA) = -\mathrm{tr}(AB) = -\mathrm{tr}(A'B)$. That is, $\mathrm{tr}(A'B) = -\mathrm{tr}(A'B) \Rightarrow \mathrm{tr}(A'B) = 0 \Rightarrow \mathrm{tr}(AB) = 0, \mathrm{tr}(AB') = 0$. □

For the next problem, let $A = A^*$ be a $p \times p$ Hermitian matrix of arbitrary elements a_{jk}. Let $\tilde{X} = \tilde{X}^*$ be a Hermitian matrix in the complex domain with distinct complex scalar variables as elements except for the Hermitian property. Let $A = A_1 + iA_2$, $\tilde{X} = X_1 + iX_2$, where A_1, A_2, X_1, X_2 are real. Then, $A_1 = A_1', X_1 = X_1'$ (symmetric), $A_2 = -A_2', X_2 = -X_2'$ (skew symmetric). Let $\tilde{u} = \mathrm{tr}(A^*\tilde{X})$. Then, we have the following result:

Theorem 7. *Let $A, \tilde{X}, A, A_1, A_2, X_1, X_2$, and $\tilde{u} = \mathrm{tr}(A^*\tilde{X})$ be as defined above. Then,*

$$\frac{\partial \tilde{u}}{\partial A} = 2\tilde{X} - \mathrm{diag}(\tilde{X})$$

where $\mathrm{diag}(\tilde{X})$ means a diagonal matrix consisting of only the diagonal elements from $\tilde{X}$.

Proof. Opening up $\tilde{u}$, we have the following:

$$\begin{aligned} \tilde{u} = \mathrm{tr}(A^*\tilde{X}) &= \mathrm{tr}\{(A_1' - iA_2')(X_1 + iX_2)\} \\ &= \mathrm{tr}\{A_1'X_1 + A_2'X_2 + i(A_1'X_2 - A_2'X_1)\} \\ &= \mathrm{tr}(A_1'X_1 + A_2'X_2) \end{aligned}$$

since $\mathrm{tr}(A_1'X_2) = 0, \mathrm{tr}(A_2'X_1) = 0$ by Lemma 1. Now, from the known result for symmetric matrices in the real case, we have

$$\frac{\partial}{\partial A_1}[\mathrm{tr}(A_1 X_1)] = 2X_1 - \mathrm{diag}(X_1). \tag{15}$$

If the (j, k)-th element, for $j \neq k$, in A_2 is $+a_{jk}$, then the (j, k)-th element in A_2' is $-a_{jk}$ and the (k, j)-th element in A_2' is a_{jk}. Hence, $\mathrm{tr}(A_2'X_2) = 2\sum_{j<k} a_{jk}x_{jk} = \sum_{j>k} a_{jk}x_{jk}$. Hence,

$$\frac{\partial}{\partial A_2}[\mathrm{tr}(A_2'X_2)] = 2X_2 \tag{16}$$

since the diagonal elements in A_2 and X_2 are already zeros. Combining (15) and (16), we have

$$\frac{\partial}{\partial A}[\mathrm{tr}(A^*\tilde{X})] = 2\tilde{X} - \mathrm{diag}(\tilde{X})$$

which corresponds to the result in the real case. □

2.4. Differentiation of a Determinant in the Complex Domain

Here, we will start with defining the derivative of a scalar quantity with respect to a matrix or we will define a matrix differential operator first. Let $X = (x_{jk})$ be an $m \times n$ real matrix and let $\tilde{X} = (\tilde{x}_{jk})$ be an $m \times n$ matrix in the complex domain. Then, we can

always write $\tilde{X} = X_1 + iX_2$, where $i = \sqrt{(-1)}$, X_1, X_2 are real $m \times n$ matrices. Then, matrix differential operators $\frac{\partial}{\partial X}$ and $\frac{\partial}{\partial \tilde{X}}$ will be defined as the following:

$$\frac{\partial}{\partial X} = \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \cdots & \frac{\partial}{\partial x_{1n}} \\ \vdots & \cdots & \vdots \\ \frac{\partial}{\partial x_{m1}} & \cdots & \frac{\partial}{\partial x_{mn}} \end{bmatrix} \text{ and } \frac{\partial}{\partial \tilde{X}} = \left(\frac{\partial}{\partial X_1} + i\frac{\partial}{\partial X_2} \right).$$

Consider a $p \times p$ nonsingular matrix $\tilde{X}$ in the complex domain and let $|\tilde{X}|$ be its determinant. Then, we have the following result:

Theorem 8. *For the $p \times p$ nonsingular matrix $\tilde{X}$ in the complex domain,*

$$\frac{\partial}{\partial \tilde{X}}[|\tilde{X}|] = 2|\tilde{X}^c|(\tilde{X}^*)^{-1} \text{ for a general } \tilde{X}$$

and

$$\frac{\partial}{\partial \tilde{X}}[|\tilde{X}|] = |\tilde{X}|[2\tilde{X}^{-1} - \text{diag}(\tilde{X}^{-1})] \text{ for } \tilde{X} = \tilde{X}^*.$$

Proof. The cofactor expansion of a determinant holds in the real and complex domains and it is the following:

$$\begin{aligned} |\tilde{X}| &= \tilde{x}_{11}\tilde{C}_{11} + ... + \tilde{x}_{1p}\tilde{C}_{1p} \\ &= \tilde{x}_{21}\tilde{C}_{21} + ... + \tilde{x}_{2p}\tilde{C}_{2p} \\ &\vdots \\ &= \tilde{x}_{p1}\tilde{C}_{p1} + ... + \tilde{x}_{pp}\tilde{C}_{pp} \end{aligned} \tag{17}$$

where $\tilde{C}_{jk}$ is the cofactor of $\tilde{x}_{jk}$ for all j and k. Let $\tilde{C} = (\tilde{C}_{jk})$ be the matrix of cofactors. For a general matrix, all $\tilde{x}_{jk}$'s are distinct and the corresponding $\tilde{C}_{jk}$'s are also distinct. Hence, in the real case, taking the partial derivative of the j-th line in (17) we have $\frac{\partial}{\partial x_{jk}}[|X|] = C_{jk}$ for all j and k. Hence, in the real case

$$\frac{\partial}{\partial X}[|X|] = C = |X'|(X')^{-1} = |X|(X')^{-1}.$$

This is a known result. But, in the complex case, the situation is different. Before we tackle the complex-domain situation, we will develop some necessary tools. It is a known result that $\tilde{X}^{-1} = \frac{1}{|\tilde{X}|}\tilde{C}'$. But when $\tilde{X} = \tilde{X}^*$ (Hermitian), the eigenvalues are real, and hence, the determinant is real. Then, $\tilde{X}^{-1}$ is Hermitian, thereby $\tilde{C}$ is also Hermitian. The following results will be helpful when we apply our matrix differential operators to scalar functions of matrices in the complex domain. For two scalar complex variables $\tilde{x}$ and $\tilde{y}$ the following can be easily verified.

$$\frac{\partial}{\partial \tilde{x}}[\tilde{x}\tilde{y}] = 0, \ \frac{\partial}{\partial \tilde{x}^c}[\tilde{x}^c\tilde{y}^c] = 0$$

$$\frac{\partial}{\partial \tilde{x}}[\tilde{x}^c\tilde{y}] = 2\tilde{y}, \ \frac{\partial}{\partial \tilde{x}}[\tilde{x}^c\tilde{y}^c] = 2\tilde{y}^c, \ \frac{\partial}{\partial \tilde{x}^c}[\tilde{x}\tilde{y}] = 2\tilde{y}. \tag{18}$$

For convenience, let us consider the term

$$\begin{aligned} \tilde{u}_{jk} &= \tilde{x}_{jk}^c\tilde{C}_{jk}^c = (x_{1jk} - ix_{2jk})(C_{1jk} - iC_{2jk}) \\ &= x_{1jk}C_{1jk} - x_{2jk}C_{2jk} - i(x_{1jk}C_{2jk} + x_{2jk}C_{1jk}) \end{aligned} \tag{19}$$

where $\tilde{x}_{jk} = x_{1jk} + ix_{2jk}, \tilde{C}_{jk} = C_{1jk} + iC_{2jk}$, with $x_{1jk}, x_{2jk}, C_{1jk}, C_{2jk}$ being real scalar quantities. Note that $\tilde{x}_{jk} = \tilde{x}_{kj}^*$ when $\tilde{X}$ is Hermitian. Hence, for example, when we differentiate with respect to $\tilde{x}_{21}$ it is equivalent to differentiating with respect to $\tilde{x}_{12}^*$ and vice versa. When $\tilde{x}_{12}^*\tilde{C}_{12}^*$ is differentiated with respect to $\tilde{x}_{12}$ it is equivalent to differentiating $\tilde{x}_{21}\tilde{C}_{21}$ with respect to $\tilde{x}_{12}$, and so on. Then,

$$\frac{\partial}{\partial x_{1jk}}[\tilde{u}_{jk}] = C_{1jk} - iC_{2jk}, \frac{\partial}{\partial x_{2jk}}[\tilde{u}_{jk}] = -C_{2jk} - iC_{1jk}, \Rightarrow \frac{\partial}{\partial \tilde{x}_{jk}}[\tilde{u}_{jk}] = 2\tilde{C}_{jk}^c. \quad (20)$$

This is the derivative of the complex conjugate of the (j,k)-th element in (18) for all j and k when $\tilde{X}$ is a general matrix with distinct scalar complex variables as elements. Then,

$$\frac{\partial}{\partial \tilde{X}}[|\tilde{X}|] = 0, \ \frac{\partial}{\partial \tilde{X}}[|\tilde{X}^c|] = 2\tilde{C}^c = 2|\tilde{X}^c|[\tilde{X}^*]^{-1}$$

$$\frac{\partial}{\partial \tilde{X}^c}[|\tilde{X}|] = 2\tilde{C} = 2|\tilde{X}|(\tilde{X}')^{-1}. \quad (21)$$

When $\tilde{X} = \tilde{X}^*$ (Hermitian) we have $|\tilde{X}| = |\tilde{X}^*| = |\tilde{X}^c|$. Then, the result in (20) holds for the k-th term in the j-th line in (18). But, when $\tilde{X}$ is Hermitian, there is one more element contributing to x_{1jk} and x_{2jk}. This is the j-th term in the k-th line of (18). Thus, the sum of the contributions coming from these two terms is the derivative of $|\tilde{X}^c|$ with respect to $\tilde{x}_{jk}$. The sum of the contributions is the following, observing that $x_{1jk} = x_{1kj}, C_{1jk} = C_{1kj}, x_{2jk} = -x_{2kj}, C_{2jk} = -C_{2kj}$:

$$x_{1jk}C_{1jk} - x_{2jk}C_{2jk} - i(x_{1jk}C_{2jk} + x_{2jk}C_{1jk})$$
$$+ x_{1kj}C_{1kj} - x_{2kj}C_{2kj} - i(x_{1kj}C_{2kj} + x_{2kj}C_{1kj})$$
$$= 2x_{1jk}C_{1jk} - 2x_{2jk}C_{2jk} - i[x_{1jk}(C_{2jk} + C_{2kj}) + C_{1jk}(x_{2jk} + x_{2kj})]$$
$$= 2x_{1jk}C_{1jk} - 2x_{2jk}C_{2jk} \text{ since } C_{2jk} + C_{2kj} = 0, x_{2jk} + x_{2kj} = 0.$$

Now,

$$\frac{\partial}{\partial \tilde{x}_{jk}}[\tilde{x}_{jk}^c\tilde{C}_{jk}^c + \tilde{x}_{kj}^c\tilde{C}_{kj}^c] = 2\tilde{C}_{jk}^c = 2\tilde{C}_{kj}$$

for all $j \neq k$. When $j = k$, the diagonal elements in $\tilde{X}$ and $\tilde{C}$ are real, and hence, the term occurs only once, and therefore,

$$\frac{\partial}{\partial \tilde{x}_{jj}}[\tilde{x}_{jj}^c\tilde{C}_{jj}^c] = \tilde{C}_{jj}^c = C_{jj}.$$

Since we are taking the partial derivative, the derivatives are the same as differentiation of $|\tilde{X}^c|$. Hence,

$$\frac{\partial}{\partial \tilde{X}}[|\tilde{X}^c|] = 2\tilde{C}^c - \text{diag}(\tilde{C}^c) = |\tilde{X}^*|[2(\tilde{X}^*)^{-1} - \text{diag}((\tilde{X}^*)^{-1})]$$
$$= |\tilde{X}|[2\tilde{X}^{-1} - \text{diag}(\tilde{X}^{-1})]$$

and then,

$$\frac{\partial}{\partial \tilde{X}}[|\tilde{X}|] = |\tilde{X}|[2\tilde{X}^{-1} - \text{diag}(\tilde{X}^{-1})].$$

Here, we have used two properties that for a $p \times p$ matrix B, $|B| = |B'|$ and $(B^c)' = B^* = B$ if B is Hermitian. Then, we have the following theorem: □

Theorem 9. *When the $p \times p$ nonsingular matrix $\tilde{X}$ in the complex domain is Hermitian, that is, $\tilde{X} = \tilde{X}^*$, then,*

$$\frac{\partial}{\partial \tilde{X}}[|\tilde{X}|] = |\tilde{X}|[2\tilde{X}^{-1} - \text{diag}(\tilde{X}^{-1})].$$

In the following sections, we will consider some applications of the results obtained in Section 2.

3. Principal Component Analysis in the Complex Domain

This is an application of the mathematical problem of optimization of a Hermitian form under a Hermitian-form constraint.

In many physical situations, variables occur in pairs, such as time and phase, and hence, the most appropriate representation of such variables is through complex variables because a scalar complex variable can be taken as a pair of real variables. Let $\tilde{x} = x_1 + ix_2$, $i = \sqrt{(-1)}, x_1, x_2$ are real scalar variables, be a scalar complex variable. Then, a statistical density associated with $\tilde{x}$ is a real-valued scalar function $f(\tilde{x})$ of $\tilde{x}$ such that $f(\tilde{x}) \geq 0$ over the entire complex plane, something like a hill on the complex plane, so that the total volume under the surface is unity, that is, $\int_{\tilde{x}} f(\tilde{x}) \mathrm{d}\tilde{x} = 1$, where $\mathrm{d}\tilde{x} = \mathrm{d}x_1 \wedge \mathrm{d}x_2$, the wedge product of the differentials $\mathrm{d}x_1$ and $\mathrm{d}x_2$, respectively. Then, the center of gravity of the hill $f(\tilde{x})$ is at $E[\tilde{x}] = \int_{\tilde{x}} \tilde{x} f(\tilde{x}) \mathrm{d}\tilde{x}$ and the square of the measure of scatter in $\tilde{x}$ is $\sigma^2 = E[(\tilde{x} - E(\tilde{x}))(\tilde{x} - E(\tilde{x}))^*]$, where $E[(\cdot)]$ means the expected value when $\tilde{x}$ is a scalar complex random variable and $f(\tilde{x})$ is its density.

If the scatter is small, then the variable $\tilde{x}$ is concentrated near the center of gravity $E[\tilde{x}]$. If σ^2 is large, then $\tilde{x}$ is spread thin far and wide, and hence, it is more or less unrecognizable. If a large number of scalar variables are being considered as possible variables to be included into a model, then the ones with the larger scatter are the important variables to be included into the model. For convenience, one can consider linear functions of such variables because linear functions also contain individual variables. A linear function is of the form $\tilde{u} = a_1^* \tilde{x}_1 + ... + a_p^* \tilde{x}_k$, where $a_1^*, ..., a_p^*$ are constants with $\tilde{x}_1, ..., \tilde{x}_p$ being scalar complex variables. For example, $a_2^* = 0 = ... = a_p^* = 0$ gives $\tilde{x}_1$, where a * indicates the conjugate transpose and for a scalar quantity y, $\tilde{y}^* = \tilde{y}^c$ only, where c in the exponent indicates the complex conjugate. We may write the linear function as $\tilde{u} = A^* \tilde{X}$, where

$$A = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}, A' = [a_1, ..., a_p] \text{ and } \tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_p \end{bmatrix}, \tilde{X}' = [\tilde{x}_1...., \tilde{x}_p]$$

where a prime indicates the transpose. The expected value of $A^* \tilde{X}$ is $E[\tilde{u}] = A^* E[\tilde{X}]$ and the variance–covariance matrix or the covariance matrix in $\tilde{X}$ is denoted by $\Sigma > O$ (Hermitian positive definite). Then, the variance of $\tilde{u}$, denoted by $\mathrm{Var}(\tilde{u})$, is given by $\mathrm{Var}(\tilde{u}) = E[(A^*(\tilde{X} - E(\tilde{X}))(\tilde{X} - E(\tilde{X}))^* A] = A^* \Sigma A$. The most important linear function is that linear function having the maximum variance. Hence, we may compute $\max_A[A^* \Sigma A]$. But, since $\Sigma > O$ we have a positive definite Hermitian form in $A^* \Sigma A$ and its maximum over A is at $+\infty$. Thus, unrestricted maximization does not make sense. Without loss of generality we may take $A^* A = 1$ because this can always be achieved for any non-null A. Then, the maximization amounts to maximizing $A^* \Sigma A$ within the unit sphere $A^* A = 1$. There will be a maximum and a minimum in this case. We may incorporate the restriction by using a Lagrangian multiplier. Consider

$$\tilde{w} = A^* \Sigma A - \lambda(A^* A - 1) \tag{22}$$

where λ is a Lagrangian multiplier. Maximization will be achieved by using the following result, which will be stated as a lemma.

Lemma 2. *Let $\tilde{Y} = Y_1 + iY_2, i = \sqrt{(-1)}, Y_1, Y_2$ are $p \times 1$ vectors with distinct real scalar variables as elements. Let $B = B^*$ be a $p \times p$ constant Hermitian matrix. Let the partial differential operator $\frac{\partial}{\partial Y}$ be defined as*

$$\frac{\partial}{\partial \tilde{Y}} = \frac{\partial}{\partial Y_1} + i \frac{\partial}{\partial Y_2}.$$

Then,

$$\frac{\partial}{\partial \tilde{Y}}[\tilde{Y}^* B \tilde{Y}] = 2B\tilde{Y}, B = B^* > O, \frac{\partial}{\partial \tilde{Y}}[\tilde{Y}^* \tilde{Y}] = 2\tilde{Y}.$$

Proof. Let $B = B^* = B_1 + iB_2$. When B is Hermitian, $B_1 = B_1', B_2 = -B_2'$, that is, B_1 is real symmetric and B_2 is real skew symmetric. Let $\tilde{Y} = Y_1 + iY_2, i = \sqrt{(-1)}, Y_1, Y_2$ are real. Then, $\tilde{Y}^* B \tilde{Y} = (Y_1' - iY_2')B(Y_1 + iY_2) = Y_1'BY_1 + Y_2'BY_2 + i(Y_1'BY_2 - Y_2'BY_1)$, where $Y_1'BY_1 = Y_1'B_1Y_1 + iY_1'B_2Y_1 = Y_1'B_1Y_1$ because $Y_1'B_2Y_1 = 0$ due to $B_2 = -B_2'$. From the real case it is well known that $\frac{\partial}{\partial Y_1}[Y_1'B_1Y_1] = 2B_1Y_1$. Similarly, $\frac{\partial}{\partial Y_2}[Y_2'BY_2] = 2B_1Y_2$. Now,

$$i\frac{\partial}{\partial Y_1}[Y_1'BY_2 - Y_2'BY_1] = i\frac{\partial}{\partial Y_1}[Y_1'B_1Y_2 - Y_2'B_1Y_1 + i(Y_1'B_2Y_2 - Y_2'B_2Y_1)]$$

$$= 0 + i^2\frac{\partial}{\partial Y_1}[Y_1'B_2Y_2 + Y_1'B_2Y_2] = i^2 2B_2Y_2 = -2B_2Y_2.$$

Similarly,

$$\frac{\partial}{\partial Y_2}[-Y_1'BY_2 - Y_2'BY_1] = -2i^2 B_2Y_1 = 2B_2Y_1$$

where we have used two properties. When $B_1 = B_1'$, that is, it is real symmetric, we have $Y_1'B_1Y_2 = Y_2'B_1Y_1$ because both are 1×1 real, and hence, each is equal to its transpose, and therefore, the difference is zero. When $B_2 = -B_2'$ we have $Y_2'B_2Y_1 = -Y_1'B_2Y_2$, because both are 1×1 real, and hence, each is equal to its transpose. Then, from the real operator operating on a real linear form the result follows. Similar results hold when differentiating with respect to the operator $\frac{\partial}{\partial Y_2}$. Thus, we have the following:

$$\frac{\partial}{\partial Y_1}[\tilde{Y}^* B \tilde{Y}] = 2B_1Y_1 - 2B_2Y_2 = 2B_1Y_1 + i^2 2B_2Y_2 \text{ and}$$

$$\frac{\partial}{\partial Y_2}[\tilde{Y}^* B \tilde{Y}] = 2B_1Y_2 + 2B_2Y_1.$$

Therefore,

$$(\frac{\partial}{\partial Y_1} + i\frac{\partial}{\partial Y_2})[\tilde{Y}^* B \tilde{Y}] = 2B_1(Y_1 + iY_2) + 2iB_2(Y_1 + iY_2) = 2B_1\tilde{Y} + 2iB_2\tilde{Y}$$

$$= 2B\tilde{Y}.$$

Hence, the result. For $B = I$, the identity matrix, the result on $\tilde{Y}^* \tilde{Y}$ follows, that is, $\frac{\partial}{\partial \tilde{Y}}(\tilde{Y}^* \tilde{Y}) = 2\tilde{Y}$. Now, by using Lemma 1 we can differentiate $\tilde{w}$ in (22). That is,

$$\frac{\partial}{\partial A}\tilde{w} = O, \frac{\partial}{\partial \lambda} = 0 \Rightarrow \Sigma A - \lambda A = O$$

$$\Rightarrow (\Sigma - \lambda I)A = O \Rightarrow |\Sigma - \lambda I| = 0 \tag{23}$$

where $|(\cdot)|$ means the determinant of the square matrix $(\cdot)$. From (23), $\Sigma A = \lambda A$ and pre-multiplying by A^* and using the fact that $A^* A = 1$, we have $A^* \Sigma A = \lambda$. Hence,

$$\max_{A^* A=1}[A^* \Sigma A] = \lambda_1 \text{ and } \min_{A^* A=1}[A^* \Sigma A] = \lambda_p \tag{24}$$

where λ_1 is the largest eigenvalue of $\Sigma = \Sigma^* > O$ and λ_p is the smallest eigenvalue of Σ. When Σ is Hermitian, all its eigenvalues are real and when it is Hermitian positive definite, all its eigenvalues are real positive also. Hence, the procedure is the following: Take the largest eigenvalue of Σ, say λ_1. Then, through (23), compute an eigenvector corresponding to λ_1, that is, solve $\Sigma A = \lambda_1 A$ for an A. Then, normalize this eigenvector through $A_1^* A_1 = 1$. That is, if an eigenvector corresponding to λ_1 is α_1, then compute $A_1 = \frac{1}{\sqrt{\alpha_1^* \alpha_1}}\alpha_1$. This A_1 is the normalized eigenvector corresponding to λ_1. Now, consider

$\tilde{u}_1 = A_1^* \tilde{X}$. This $\tilde{u}_1$ is the first principal component in the sense of the linear function having the maximum variance. Now, take the second largest eigenvalue λ_2. Go through the same procedure and construct the normalized eigenvector A_2 corresponding to λ_2. Then, $\tilde{u}_2 = A_2^* \tilde{X}$ is the second principal component. Continue the process and stop the process when the variance of $\tilde{u}_j$, namely, λ_j, falls below a preassigned number. If there is no preassigned number, then $\tilde{u}_1, ..., \tilde{u}_p$ will be the p principal components. Here, we have assumed that the eigenvalues of Σ are distinct. When the eigenvalues are distinct, we can show that the eigenvectors corresponding to the distinct eigenvalues of a symmetric or Hermitian matrix are orthogonal to each other. Hence, our principal components will be orthogonal to each other in the sense that the joint dispersion in the pair $(\tilde{u}_i, \tilde{u}_j)$ is 0 for $i \neq j$ or the covariance between $\tilde{u}_i$ and $\tilde{u}_j$ is zero when $i \neq j$. The covariance is defined as the following: Let $\tilde{U}$ be a $p \times 1$ vector in the complex domain and let $\tilde{V}$ be a $q \times 1$ vector in the complex domain. Then, the covariance of $\tilde{U}$ on $\tilde{V}$ is defined and denoted as $\mathrm{Cov}(\tilde{U}, \tilde{V}) = E[(\tilde{U} - E(\tilde{U}))(\tilde{V} - E(\tilde{V}))^*]$, whenever this expected value exists, so that when $\tilde{V} = \tilde{U}$, then $\mathrm{Cov}(\tilde{U}, \tilde{V}) = \mathrm{Cov}(\tilde{U})$, the covariance matrix in $\tilde{U}$, and when $p = 1$, it is the variance of the scalar complex variable $\tilde{u}$. $\square$

When the covariance matrix Σ is unknown, then we may construct sample principal components. Let our population be the $p \times 1$ vector $\tilde{X}$, $\tilde{X}' = [\tilde{x}_1, ..., \tilde{x}_p]$, where $\tilde{x}_j, j = 1, ..., p$ are distinct scalar complex variables. Consider n independently and identically distributed (iid) such p-vectors. Then, we have a simple random sample of size n from $\tilde{X}$. Then, the sample matrix is the $p \times n, n > p$ matrix, denoted as the following:

$$\tilde{\mathbf{X}} = [\tilde{X}_1, ..., \tilde{X}_n] = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & ... & \tilde{x}_{1n} \\ \tilde{x}_{21} & \tilde{x}_{22} & ... & \tilde{x}_{2n} \\ \vdots & \vdots & ... & \vdots \\ \tilde{x}_{p1} & \tilde{x}_{p2} & ... & \tilde{x}_{pn} \end{bmatrix}.$$

Let the sample average be denoted by $\bar{\tilde{X}} = \frac{1}{n}[\tilde{X}_1 + ... + \tilde{X}_n]$ and the matrix of sample averages be denoted by the bold letter $\bar{\tilde{\mathbf{X}}} = [\bar{\tilde{X}}, ..., \bar{\tilde{X}}]$. Then, the sample sum of products matrix $\tilde{S}$ is given by

$$\tilde{S} = [\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}}][\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}}]^* = (\tilde{s}_{jk})$$

where $\tilde{s}_{jk} = \sum_{r=1}^{n}(\tilde{x}_{jr} - \bar{\tilde{x}}_j)(\tilde{x}_{rk} - \bar{\tilde{x}}_k)^*$. The motivation in using the sample sum of products matrix $\tilde{S}$ is that $\frac{1}{n-1}\tilde{S}$ is an unbiased estimator of Σ. Since we will be normalizing the eigenvectors, operate with $\tilde{S}$ itself. Compute the eigenvalues of $\tilde{S}$. Take the largest eigenvalue of $\tilde{S}$. Call it m_1. Construct an eigenvector corresponding to m_1 and normalize it through $M_1^* M_1 = 1$, where M_1 is the normalized eigenvector corresponding to m_1. Then, $\tilde{v}_1 = M_1^* \tilde{X}$ is the first sample principal component. When the columns of the sample matrix are not linearly related then we have $\tilde{S} = \tilde{S}^* > O$ (Hermitian positive definite) and all eigenvalues $m_1, ..., m_p$ will be positive. We assume that the eigenvalues are distinct $m_1 > m_2 > ... > m_p$. This will be true almost surely. Now, take m_2 and construct M_2 and the second principal component $\tilde{v}_2 = M_2^* \tilde{X}$ and continue the process. We can show that the covariances between $\tilde{v}_j$ and $\tilde{v}_k$ will be zeros for all $j \neq k$. This property follows from the fact that when the matrix is symmetric or Hermitian, the eigenvectors corresponding to distinct eigenvalues are orthogonal. When the population $\tilde{X}$ is p-variate complex Gaussian, then we can show that $\tilde{S}$ will be a complex Wishart distributed with degrees of freedom $n - 1$ and parameter matrix Σ. The distributions of the largest, smallest, and j-th largest eigenvalues and the corresponding eigenvectors of $\tilde{S}$ in the complex domain are given in [8].

4. Canonical Correlation Analysis in the Complex Domain

This is an application of the mathematical problem of optimization of a bilinear form in the complex domain, under two Hermitian-form constraints. The following application is regarding the prediction of one set of variables by using another set of variables.

Consider two sets of scalar complex variables

$$S_1 = \{\tilde{x}_1, ..., \tilde{x}_p\} \text{ and } S_2 = \{\tilde{y}_1, ..., \tilde{y}_q\}$$

where p need not be equal to q. Consider the appended vector

$$\tilde{Z} = \begin{bmatrix} \tilde{X} \\ \tilde{Y} \end{bmatrix}, \tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_p \end{bmatrix}, \tilde{Y} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_q \end{bmatrix}, E[\tilde{X}] = \mu_x, E[\tilde{Y}] = \mu_y,$$

$$\Sigma = E\left[\begin{bmatrix} \tilde{X} - \mu_x \\ \tilde{Y} - \mu_y \end{bmatrix} [\tilde{X}^* - \mu_x^*, \tilde{Y}^* - \mu_y^*] \right] = \text{Cov}(\tilde{Z}) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where $\Sigma_{11} = \text{Cov}(\tilde{X}), \Sigma_{22} = \text{Cov}(\tilde{Y}), \Sigma_{12} = \text{Cov}(\tilde{X}, \tilde{Y})$, and $\Sigma_{12}^* = \Sigma_{21}$, and vice-versa. That is, Σ is the covariance matrix in $\tilde{Z}$, Σ_{11} is the covariance matrix in $\tilde{X}$, Σ_{22} is the covariance matrix in $\tilde{Y}$, and so on. Also, $\Sigma = \Sigma^* > O, \Sigma_{11} = \Sigma_{11}^* > O, \Sigma_{22} = \Sigma_{22}^* > O$. Our aim is to predict the variables in the set S_1 by using the variables in the set S_2 and vice-versa, and obtain the "best" predictors; "best" in the sense of having the maximum joint dispersion. In order to represent each set S_1 and S_2, we will take arbitrary linear functions of the variables in each set. Consider linear functions $\tilde{u} = A^*\tilde{X}$ and $\tilde{v} = B^*\tilde{Y}$, where

$$A = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}, B = \begin{bmatrix} b_1 \\ \vdots \\ b_q \end{bmatrix}, \begin{matrix} \tilde{u} &= A^*\tilde{X} = a_1^*\tilde{x}_1 + ... + a_p^*\tilde{x}_p \\ \tilde{v} &= B^*\tilde{Y} = b_1^*\tilde{y}_1 + ... + b_q^*\tilde{y}_q \end{matrix},$$

where $\tilde{X}$ and $\tilde{Y}$ are listed above already. Since a_j and b_j are scalar constant quantities, $a_j^* = a_j^c, j = 1, ..., p$ and $b_j^* = b_j^c, j = 1, ..., q$. Variances for linear functions are already seen in Section 2. Therefore, $\text{Var}(\tilde{u}) = A^*\Sigma_{11}A, \text{Var}(\tilde{v}) = B^*\Sigma_{22}B$ and $\text{Cov}(\tilde{u}, \tilde{v}) = A^*\Sigma_{12}B$, $\Sigma_{21} = \Sigma_{12}^*, B^*\Sigma_{21}A = \text{Cov}(\tilde{v}, \tilde{u})$. Here, Σ_{12} and Σ_{21} can be taken as measures of joint dispersion or joint variation between $\tilde{X}$ and $\tilde{Y}$ and $A^*\Sigma_{12}B = \text{Cov}(\tilde{u}, \tilde{v})$ as the joint dispersion between $\tilde{u}$ and $\tilde{v}$. As a criterion for the "best" predictor of $\tilde{u}$ by using $\tilde{v}$ and vice versa, we may take that the pair $\tilde{u}$ and $\tilde{v}$ have the maximum joint variation. The best predictor of $\tilde{u}$ by using $\tilde{v}$ as that pair having the maximum $A^*\Sigma_{12}B$ and the best predictor of $\tilde{v}$ by using $\tilde{u}$ as that pair having the maximum value for $B^*\Sigma_{21}A$. Since covariances depend upon the units of measurements of the variables involved, we may take a scale-free covariance by taking

$$\rho = \frac{\text{Cov}(\tilde{u}, \tilde{v})}{\sqrt{\text{Var}(\tilde{u})\text{Var}(\tilde{v})}} = \frac{A^*\Sigma_{12}B}{\sqrt{[A^*\Sigma_{11}A][B^*\Sigma_{22}B]}}.$$

Further, as explained in Section 2, without loss of generality we may take $A^*\Sigma_{11}A = 1$ and $B^*\Sigma_{22}B = 1$ or confine the bilinear form (hyperboloid) within unit positive definite Hermitian forms (ellipsoids) in order to prevent them from going to $+\infty$. Hence, our procedure simplifies to optimizing $A^*\Sigma_{12}B$ subject to the conditions $A^*\Sigma_{11}A = 1, B^*\Sigma_{22}B = 1$ and computing that pair of A and B which will maximize $A^*\Sigma_{12}B$. As before, we may use the Lagrangian multipliers λ_1 and λ_2 and consider the function

$$\tilde{w} = A^*\Sigma_{12}B - \lambda_1(A^*\Sigma_{11}A - 1) - \lambda_2(B^*\Sigma_{22}B - 1). \tag{25}$$

In order to optimize this $\tilde{w}$ we need one result on differentiation of a bilinear form, which will be stated as a lemma.

Lemma 3. *Let $\tilde{X} = X_1 + iX_2, i = \sqrt{(-1)}, X_1, X_2$ are real, $p \times 1$ vectors of distinct real scalar variables x_{1j} and x_{2j}, respectively, where $\tilde{x}_j = x_{j1} + ix_{j2}, j = 1, ..., p$. Let $\tilde{Y} = Y_1 + iY_2$, $i = \sqrt{(-1)}, Y_1, Y_2$ are real, $q \times 1$ vectors of distinct real scalar variables y_{j1} and y_{j2}, respec-*

tively, where $\tilde{y}_j = y_{j1} + iy_{j2}, j = 1, ..., q$, where $x_{j1}, x_{j2}, y_{j1}, y_{j2}$ are real. Let the partial differential operators be as defined in Section 2, namely,

$$\frac{\partial}{\partial X_1} = \begin{bmatrix} \frac{\partial}{\partial x_{11}} \\ \vdots \\ \frac{\partial}{\partial x_{p1}} \end{bmatrix}, \frac{\partial}{\partial X_2} = \begin{bmatrix} \frac{\partial}{\partial x_{12}} \\ \vdots \\ \frac{\partial}{\partial x_{p2}} \end{bmatrix}, \frac{\partial}{\partial \tilde{X}} = (\frac{\partial}{\partial X_1} + i\frac{\partial}{\partial X_2})$$

and similar operators involving $\tilde{Y} = Y_1 + iY_2$. Then,

$$\frac{\partial}{\partial \tilde{X}}[\tilde{X}^* A\tilde{Y}] = 2A\tilde{Y} \text{ and } \frac{\partial}{\partial \tilde{Y}}[\tilde{Y}^* A^* \tilde{X}] = 2A^* \tilde{X}.$$

Proof. Opening up $\tilde{X}^* A\tilde{Y}$ we have the following:

$$\tilde{X}^* A\tilde{Y} = (X_1' - iX_2')A(Y_1 + iY_2) = X_1'AY_1 + X_2'AY_2 + i(X_1'AY_2 - X_2'AY_1).$$

Then, from the known results in the real case, we have the following:

$$\frac{\partial}{\partial X_1}[\tilde{X}^* A\tilde{Y}] = AY_1 + iAY_2 \text{ and } \frac{\partial}{\partial X_2}[\tilde{X}^* A\tilde{Y}] = AY_2 - iAY_1$$

irrespective of whether A is real or in the complex domain. Then,

$$\frac{\partial}{\partial \tilde{X}}[\tilde{X}^* A\tilde{Y}] = (\frac{\partial}{\partial X_1} + i\frac{\partial}{\partial X_2})[\tilde{X}^* A\tilde{Y}] = 2A\tilde{Y}.$$

Similarly, $\frac{\partial}{\partial \tilde{Y}}[\tilde{Y}^* A^* \tilde{X}] = 2A^* \tilde{X}$. This completes the proof. $\square$

Now, differentiating $\tilde{w}$ in (25), we have the following:

$$\frac{\partial \tilde{w}}{\partial A} = O \Rightarrow \Sigma_{12}B - \lambda_1 \Sigma_{11}A = O \tag{26}$$

$$\frac{\partial \tilde{w}^*}{\partial B} = O \Rightarrow \Sigma_{21}A - \lambda_2 \Sigma_{22}B = O. \tag{27}$$

Now, premultiply (26) by A^* and (27) by B^* to obtain the following, observing that $\Sigma_{21} = \Sigma_{12}^*$:

$$A^* \Sigma_{12}B = \lambda_1, B^* \Sigma_{21}A = \lambda_2, \Rightarrow \lambda_2 = \lambda_1^c$$

or $\lambda_1 = \lambda, \lambda_2 = \lambda^c$. Take B from (27) and substitute in (26) to obtain the following:

$$(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda\lambda^c I)A = O, \lambda\lambda^c = |\lambda|^2. \tag{28}$$

This shows that $\lambda\lambda^c = |\lambda|^2 = \mu$ is an eigenvalue of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, where the matrix is $p \times p$. From symmetry, it follows that μ is also an eigenvalue of $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, where the matrix is $q \times q$. Hence, all the nonzero μ's are common to both of these matrices. Hence, the procedure is the following: If $p \leq q$, then compute the eigenvalues of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, otherwise, compute the eigenvalues of the other matrix $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, both will give the same nonzero eigenvalues. Let μ_1 be the largest and μ_r be the smallest nonzero eigenvalues. Then, we have the results

$$\max_{A^* \Sigma_{11}A=1, B^* \Sigma_{22}B=1}[A^* \Sigma_{12}B] = \mu_1$$

and

$$\min_{A^* \Sigma_{11}A=1, B^* \Sigma_{22}B=1}[A^* \Sigma_{12}B] = \mu_r.$$

Then, the procedure is the following: If $p \leq q$, then compute all the eigenvalues of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Let the largest eigenvalue be μ_1. Then, compute one eigenvector corresponding to μ_1. Use the equation $(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \mu_1\Sigma_{11})A = O$. Let it be A_{11}. Then, normalize it through $A_{11}^*\Sigma_{11}A_{11} = 1$. That is, compute $A_1 = \frac{1}{\sqrt{A_{11}^*\Sigma_{11}A_{11}}}A_{11}$. Then, compute $\tilde{u}_1 = A_1^*\tilde{X}$. Then, use the same eigenvalue μ_1 and compute one eigenvector from the equation $(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \mu_1\Sigma_{22})B = O$. Let it be B_{11}. Then, normalize it through $B_{11}^*\Sigma_{22}B_{11} = 1$, that is, compute $B_1 = \frac{1}{\sqrt{B_{11}^*\Sigma_{22}B_{11}}}B_{11}$. Now, compute $\tilde{v}_1 = B_1^*\tilde{Y}$. Then, $(\tilde{u}_1, \tilde{v}_1)$ is the first pair of canonical variables in the sense $\tilde{u}_1$ is the best predictor of $\tilde{v}_1$ and vice-versa. Now, take the second largest eigenvalue μ_2 of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Then, compute one eigenvector corresponding to μ_2, that is, solve the equation $(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \mu_2\Sigma_{11})A = O$. Let A_{21} be that eigenvector. Normalize it, that is, compute $A_2 = \frac{1}{\sqrt{A_{21}^*\Sigma_{11}A_{21}}}A_{21}$. Now, compute $\tilde{u}_2 = A_2^*\tilde{X}$. Use the same μ_2 and solve for B from the equation $(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \mu_2\Sigma_{22})B = O$. Let B_{21} be one solution. Then, normalize it, that is, compute $B_2 = \frac{1}{\sqrt{B_{21}^*\Sigma_{22}B_{21}}}B_{21}$. Now, compute $\tilde{v}_2 = B_2^*\tilde{Y}$. Then, $(\tilde{u}_2, \tilde{v}_2)$ is the second pair of canonical variables. Continue the process until μ_j falls below a preassigned limit. If there is no such preassigned limit, then compute all the pairs, that is, p if $p \leq q$; otherwise q. If $q < p$, then start with the computation of the eigenvalues of $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and proceed parallel to the steps used in the case $p \leq q$. Observe that the symmetric format of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is $\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$. This form is also available from the same starting equation (28). The symmetric format can always be written in the form C^*C for some matrix C, and hence, the symmetric form is either Hermitian positive definite or Hermitian positive semi-definite, and therefore, all the nonzero eigenvalues are positive. Let us assume that the eigenvalues $\mu_1, \mu_2, \ldots$ are distinct, $\mu_1 > \mu_2 > \ldots > \mu_p$ if $p \leq q$. It is a known result that eigenvectors corresponding to distinct eigenvalues of Hermitian or symmetric matrices are orthogonal to each other. Hence, $\tilde{u}_1, \tilde{u}_2, \ldots$ are non-correlated. Similarly, $\tilde{v}_1, \tilde{v}_2, \ldots$ are non-correlated.

If $\Sigma_{11}, \Sigma_{12}, \Sigma_{22}$ are not available, then take a simple random sample of size n, $n > p + q$, from $\tilde{Z} = \begin{bmatrix} \tilde{X} \\ \tilde{Y} \end{bmatrix}$. Then, compute the sample sum of products matrix:

$$\tilde{S} = \begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{bmatrix},$$

as for the case of principal component analysis. Now, continue with $\tilde{S}_{jk}$'s as with Σ_{jk}'s. Then, we will obtain the pairs of sample canonical variables. Some distributional aspects of sample canonical variables and the sample canonical correlation matrix $\tilde{U} = \tilde{S}_{11}^{-\frac{1}{2}}\tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21}\tilde{S}_{11}^{-\frac{1}{2}}$ are discussed in [8]. Note that when S_{11} is 1×1, then we have the square of the multiple correlation coefficient in $\tilde{U} = \tilde{u}$, which is scalar. In this case, our starting set S_1 will have only one complex scalar variable and the set S_2 will have q variables. Here, the problem is to predict one variable in S_1 by using the variables in S_2. The exact distributions of the canonical correlation matrix $\tilde{U}$ and the square of the absolute value of the multiple correlation coefficient $\tilde{u}$ are available in explicit forms in [8].

5. Covariance and Correlation in the Complex Domain

Consider scalar complex variables first. Let $\tilde{x}_1$ and $\tilde{y}_1$ be two scalar complex variables or scalar variables in the complex domain. Then, the mean values or expected values of $\tilde{x}_1$ and $\tilde{y}_1$ are, respectively, $E[\tilde{x}_1] = \int_{\tilde{x}_1} \tilde{x}_1 f(\tilde{x}_1)\mathrm{d}\tilde{x}_1$, $E[\tilde{y}_1] = \int_{\tilde{y}_1} \tilde{y}_1 g(\tilde{y}_1)\mathrm{d}\tilde{y}_1$, where $E[\cdot]$ means the expected value of $[\cdot]$, and f and g are the densities of $\tilde{x}_1$ and $\tilde{y}_1$, respectively. Let $\tilde{x} = \tilde{x}_1 - E[\tilde{x}_1], \tilde{y} = \tilde{y}_1 - E[\tilde{y}_1]$. Let $\tilde{x} = x_{11} + ix_{12}, i = \sqrt{(-1)}, x_{11}, x_{12}$ are real scalar variables, and let $\tilde{y} = y_{11} + iy_{12}, i = \sqrt{(-1)}, y_{11}, y_{12}$ are real scalar variables. Then, $E[\tilde{x}] = 0$, $E[\tilde{y}] = 0$. Then, the variance, denoted by $\mathrm{Var}(\cdot)$, and covariance, denoted by $\mathrm{Cov}(\cdot, \cdot)$, are the following:

$$\text{Var}(\tilde{x}_1) = \text{Var}(\tilde{x}) = E[\tilde{x}\tilde{x}^*] = \sigma_{11}$$
$$\text{Var}(\tilde{y}_1) = \text{Var}(\tilde{y}) = E[\tilde{y}\tilde{y}^*] = \sigma_{22}$$
$$\text{Cov}(\tilde{x}_1, \tilde{y}_1) = \text{Cov}(\tilde{x}, \tilde{y}) = E[\tilde{x}\tilde{y}^*] = \sigma_{12}$$

where, for example, $\tilde{x}^* = \tilde{x}^c$, with $*$ indicating the conjugate transpose and c indicating the conjugate only. Here, we have only scalar variables, and hence, the complex conjugate transpose is only complex conjugate. For convenience, we have used the notation $\sigma_{11}, \sigma_{22}, \sigma_{12}$. Then, $\sigma_{21} = E[\tilde{y}\tilde{x}^*] = E[\tilde{y}\tilde{x}^c] = \text{Cov}(\tilde{y}, \tilde{x})$. For a scalar complex variable $\tilde{x} = x_{11} + ix_{12}$, $\tilde{x}^* = x'_{11} - ix'_{12} = x_{11} - ix_{12} = \tilde{x}^c$. Let us examine the variances of sum and difference. Let

$$\tilde{u} = \frac{\tilde{x}_1 - E[\tilde{x}_1]}{\sqrt{Var(\tilde{x}_1)}} = \frac{\tilde{x}}{\sqrt{\sigma_{11}}}, \tilde{v} = \frac{\tilde{y}_1 - E[\tilde{y}_1]}{\sqrt{Var(\tilde{y}_1)}} = \frac{\tilde{y}}{\sqrt{\sigma_{22}}}.$$

Then,

$$\text{Var}(\tilde{u} + \tilde{v}) = E[(\tilde{u} + \tilde{v} - E(\tilde{u} + \tilde{v}))(\tilde{u} + \tilde{v} - E(\tilde{u} + \tilde{v}))^*]$$
$$= E\left[\frac{\tilde{x}\tilde{x}^*}{\sigma_{11}} + \frac{\tilde{y}\tilde{y}^*}{\sigma_{22}} + \frac{\tilde{x}\tilde{y}^* + \tilde{y}\tilde{x}^*}{\sqrt{\sigma_{11}\sigma_{22}}}\right]$$

This can be simplified as the following:

$$\text{Var}(\tilde{u} + \tilde{v}) = 2 + 2\frac{[\text{Cov}(\tilde{x}, \tilde{y}) + \text{Cov}(\tilde{y}, \tilde{x})]}{2\sqrt{\sigma_{11}\sigma_{22}}} \geq 0 \tag{29}$$

$$\text{Var}(\tilde{u} - \tilde{v}) = \frac{E[\tilde{x}\tilde{x}^*]}{\sigma_{11}} + \frac{E[\tilde{y}\tilde{y}^*]}{\sigma_{22}} - \frac{E[\tilde{x}\tilde{y}^* + \tilde{y}\tilde{x}^*]}{\sqrt{\sigma_{11}\sigma_{22}}}$$
$$= 2 - 2\frac{[\text{Cov}(\tilde{x}, \tilde{y}) + \text{Cov}(\tilde{y}, \tilde{x})]}{2\sqrt{\sigma_{11}\sigma_{22}}} \geq 0 \tag{30}$$

From (29) and (30) we have

$$-1 \leq \frac{\text{Cov}(\tilde{x}, \tilde{y}) + \text{Cov}(\tilde{y}, \tilde{x})}{2\sqrt{\sigma_{11}\sigma_{22}}} \leq 1. \tag{31}$$

Let us examine the quantity, $\text{Cov}(\tilde{x}, \tilde{y}) + \text{Cov}(\tilde{y}, \tilde{x}) = E[\tilde{x}\tilde{y}^*] + E[\tilde{y}\tilde{x}^*]$. Note that

$$E[\tilde{x}\tilde{y}^*] = E[(x_{11} + ix_{12})(y_{11} - iy_{12})] = E[x_{11}y_{11} + x_{12}y_{12} + i(x_{12}y_{11} - x_{11}y_{12})]$$
$$E[\tilde{y}\tilde{x}^*] = E[x_{11}y_{11} + x_{12}y_{12} + i(y_{12}x_{11} - y_{11}x_{12})] \Rightarrow$$
$$E[\tilde{x}\tilde{y}^*] + E[\tilde{y}\tilde{x}^*] = \text{Cov}(\tilde{x}, \tilde{y}) + \text{Cov}(\tilde{y}, \tilde{x}) = 2E[x_{11}y_{11} + x_{12}y_{12}] = 2E[\Re(\tilde{x}\tilde{y}^*)]$$

Hence,

$$\frac{1}{2}\left[\frac{\text{Cov}(\tilde{x}, \tilde{y}) + \text{Cov}(\tilde{y}, \tilde{x})}{\sqrt{\sigma_{11}\sigma_{22}}}\right] = E\left[\frac{x_{11}y_{11} + x_{12}y_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}\right]$$
$$= E\left[\frac{\Re(\tilde{x}\tilde{y}^*)}{\sqrt{\sigma_{11}\sigma_{22}}}\right].$$

Therefore, we may define the correlation coefficient in the complex domain, denoted by $\tilde{r}$, as the following:

$$\tilde{r} = \frac{\Re(\text{Cov}(\tilde{x}, \tilde{y}))}{\sqrt{\sigma_{11}\sigma_{22}}} \Rightarrow -1 \leq \tilde{r} \leq 1 \tag{32}$$

where $\Re(\cdot)$ denotes the real part in $(\cdot)$. Also, note that $|\tilde{x}\tilde{y}^*|^2 = |\tilde{x}|^2|\tilde{y}|^2$. This will motivate us to examine the dot product of two vectors in the complex domain and the Cauchy–Schwarz inequality in the complex domain.

5.1. The Cauchy–Schwarz Inequality in the Complex Domain

Let $\tilde{X}, \tilde{X}' = [\tilde{x}_1, ..., \tilde{x}_p]$, and $\tilde{Y}, \tilde{Y}' = [\tilde{y}_1, ..., \tilde{y}_p]$ be two $p \times 1$ vectors in the complex domain. We will define the dot product between $\tilde{X}$ and $\tilde{Y}$, and it will be denoted and defined as the following: $\tilde{X} \odot \tilde{Y} = \tilde{X}^*\tilde{Y}$ and

$$
\begin{aligned}
|\tilde{X} \odot \tilde{Y}|^2 = |\tilde{X}^*\tilde{Y}|^2 &= |\tilde{x}_1^c \tilde{y}_1|^2 + ... + |\tilde{x}_p^c \tilde{y}_p|^2 \\
&= |\tilde{x}_1|^2|\tilde{y}_1|^2 + ... + |\tilde{x}_p|^2|\tilde{y}_p|^2 \\
&= (x_{11}^2 + x_{12}^2)(y_{11}^2 + y_{12}^2) + ... + (x_{p1}^2 + x_{p2}^2)(y_{p1}^2 + y_{p2}^2) \\
&\leq (x_{11}^2 + x_{12}^2 + ... + x_{p1}^2 + x_{p2}^2)(y_{11}^2 + y_{12}^2 + ... + y_{p1}^2 + y_{p2}^2) \\
&\Rightarrow |\tilde{X} \odot \tilde{Y}| \leq |\tilde{X}|\,|\tilde{Y}|.
\end{aligned}
\tag{33}
$$

Thus, the Cauchy–Schwarz inequality holds for the complex domain also.

5.2. Minimum-Variance Unbiased Estimators in the Complex Domain

In the class of linear estimators $A^*\tilde{X}$ for the parametric function $g(\theta)$, which linear function is the minimum-variance unbiased estimator for $g(\theta)$? Let

$$
A = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}, \; \tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_p \end{bmatrix} \Rightarrow E[A^*\tilde{X}] = A^*E[\tilde{X}] = A^*\tilde{\mu} = g(\theta),
$$

and $\text{Var}(A^*\tilde{X}) = A^*\Sigma A, E[\tilde{X}] = \tilde{\mu}$ where Σ is the covariance matrix in $\tilde{X}$. Our aim here is to minimize $A^*\Sigma A$ subject to the constraint $A^*\tilde{\mu} = g(\theta)$ (given). Let λ be a Lagrangian multiplier and let $\tilde{w} = A^*\Sigma A - \lambda(A^*\tilde{\mu} - g(\theta))$. Then,

$$
\frac{\partial \tilde{w}}{\partial A} = O \Rightarrow 2\Sigma A - 2\lambda\tilde{\mu} = O \Rightarrow \Sigma A = \lambda\tilde{\mu}.
\tag{34}
$$

That is, $A^*\Sigma A = \lambda A^*\tilde{\mu} = \lambda g(\theta)$. Hence, $g(\theta)$ multiplied by the minimum value of λ gives the minimum of $A^*\Sigma A$. From (34),

$$
A = \lambda\Sigma^{-1}\tilde{\mu} \Rightarrow g(\theta) = A^*\tilde{\mu} = \lambda^c\tilde{\mu}^*\Sigma^{-1}\tilde{\mu} \Rightarrow \lambda^c = \frac{g(\theta)}{\tilde{\mu}^*\Sigma^{-1}\tilde{\mu}}.
$$

Then,

$$
\lambda g(\theta) = \frac{gg^*}{\tilde{\mu}^*\Sigma^{-1}\tilde{\mu}} = \frac{|g|^2}{\tilde{\mu}^*\Sigma^{-1}\tilde{\mu}}
$$

is the minimum value of the variance of our linear estimator. Hence, the minimum-variance unbiased estimator is

$$
T(\tilde{X}) = \frac{g}{\tilde{\mu}^*\Sigma^{-1}\tilde{\mu}}[\tilde{\mu}^*\Sigma^{-1}\tilde{X}].
$$

Note also that

$$
g = A^*\tilde{\mu} = (A^*\Sigma^{\frac{1}{2}})(\Sigma^{-\frac{1}{2}}\tilde{\mu}) \leq \sqrt{A^*\Sigma A}\sqrt{\tilde{\mu}\Sigma^{-1}\tilde{\mu}} \Rightarrow \sqrt{A^*\Sigma A} \geq \frac{g}{\tilde{\mu}^*\Sigma^{-1}\tilde{\mu}}.
$$

The first inequality follows from the inequality established in Section 5.1.

5.3. Cramer–Rao-Type Inequality in the Complex Domain

Let $\tilde{x}_1, ..., \tilde{x}_n$ be a simple random sample from the population designated by the density $f(\tilde{x}_j)$, where $\tilde{x}_j, j = 1, ..., n$ are independently and identically distributed (iid). Then, the joint density $L = \prod_{j=1}^{n} f(\tilde{x}_j)$. Let $T(\tilde{x}_1, ..., \tilde{x}_n)$, denoted as $T(\tilde{X})$, be a statistic with the expected value $g(\theta)$. That is, $E[T] = \int_{\tilde{X}} TL \, d\tilde{X} = g(\theta)$. Differentiating with respect to θ, we have

$$\frac{\partial g(\theta)}{\partial \theta} = \int_{\tilde{X}} T(\frac{\partial L}{\partial \theta}) d\tilde{X} = \int_{\tilde{X}} T(\frac{\partial \ln L}{\partial \theta}) L \, d\tilde{X} = E[T(\frac{\partial \ln L}{\partial \theta})] \tag{35}$$

where we have assumed that the support of $\tilde{X}$ is free of θ and differentiation inside the integral is valid. But, from the total integral being one, we have

$$\int_{\tilde{X}} L \, d\tilde{X} = 1 \Rightarrow \int_{\tilde{X}} (\frac{\partial \ln L}{\partial \theta}) L \, d\tilde{X} = 0 \Rightarrow E[\frac{\partial \ln L}{\partial \theta}] = 0. \tag{36}$$

From (35) and (36), we have $E[T(\frac{\partial \ln L}{\partial \theta})] = \text{Cov}(T, \frac{\partial \ln L}{\partial \theta})$ because for any two scalar random variables u and v real, or $\tilde{u}$ and $\tilde{v}$ in the complex domain, $\text{Cov}(u, v) = E[(u - E(u))(v - E(v))] = E[u(v - E(v))] = E[(u - E(u))v]$ and the corresponding results in the complex domain also hold. Therefore, since $E[\frac{\partial \ln L}{\partial \theta}] = 0$, $E[T(\frac{\partial \ln L}{\partial \theta})] = \text{Cov}(T, \frac{\partial \ln L}{\partial \theta})$. Then,

$$|\text{Cov}(T, \frac{\partial \ln L}{\partial \theta})| \leq \sqrt{\text{Var}(T)}\sqrt{\text{Var}(\frac{\partial \ln L}{\partial \theta})} \Rightarrow \text{Var}(\tilde{T}) \geq \frac{|\frac{\partial g}{\partial \theta}|^2}{\text{Var}(\frac{\partial \ln L}{\partial \theta})}. \tag{37}$$

Note that

$$\text{Var}(\frac{\partial \ln L}{\partial \theta}) = E[(\frac{\partial \ln L}{\partial \theta})^2] = n\text{Var}(\frac{\partial \ln f}{\partial \theta}) = nE[(\frac{\partial \ln L}{\partial \theta})^2]. \tag{38}$$

This shows that the Cramer–Rao-type inequality holds in the complex domain also.

5.4. Least Square Estimation in Linear Models in the Complex Domain

Let us examine whether the least square procedure, in the class of linear models, holds in the complex domain also. Let $\tilde{x}_1, ..., \tilde{x}_k$ be preassigned complex numbers or observations on k random variables in the complex domain. Let $\tilde{y}$ be a scalar complex variable. Then, a linear model of $\tilde{x}_1, ..., \tilde{x}_k$ for predicting $\tilde{y}$, can be of the following form:

$$\tilde{y} = a_0^c + \beta^* \tilde{X}, \beta = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}, \tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_k \end{bmatrix}.$$

But, in order to predict $\tilde{y}$ by using linear predictors we must know the conditional distribution of $\tilde{y}$, given $\tilde{x}_1, ..., \tilde{x}_k$, and also the conditional expectation must be linear in $\tilde{x}_1, ..., \tilde{x}_k$. If the conditional distribution is not known, then we may use a distribution-free procedure. One such procedure is the estimation procedure by using the least square method. In this method, we set up a corresponding model of the following form for the j-th observation on $\tilde{y}$, namely, $\tilde{y}_j$, for $j = 1, ..., n$, where $n > k + 1$ is the sample size.

$$\tilde{y} = a_0 + a_1 \tilde{x}_{1j} + ... + a_k \tilde{x}_{kj} + \epsilon_j, j = 1, ..., n \tag{39}$$

corresponding to the linear model in the real case, where ϵ_j is the random part or the sum total contributions coming from unknown factors, corresponding to $\tilde{y}_j$. Then, if we sum up the observations and divide by n we obtain the sample averages $\bar{\tilde{y}}, \bar{\tilde{x}}_r, r = 1, ..., k$, where, for example, $\bar{\tilde{x}}_r = \frac{1}{n}\sum_{j=1}^{n} \tilde{x}_{rj}$. Then, from (39), we have

$$\bar{\tilde{y}} = a_0 + a_1 \bar{\tilde{x}}_1 + ... + a_k \bar{\tilde{x}}_k + 0 \Rightarrow a_0 = \bar{\tilde{y}} - a_1 \bar{\tilde{x}}_1 - ... - a_k \bar{\tilde{x}}_k$$
$$\Rightarrow c_0^c = \bar{\tilde{y}} - a_1^c \bar{\tilde{x}}_1 - ... - a_k^c \bar{\tilde{x}}_k. \tag{40}$$

We have taken the error sum as zero without much loss of generality. Since a_0 is available from (40), we may rewrite (39) as follows:

$$\tilde{y}_j - \bar{\tilde{y}} = a_1(\tilde{x}_{1j} - \bar{\tilde{x}}_1) + ... + a_k(\tilde{x}_{kj} - \bar{\tilde{x}}_k) + \epsilon_j - \bar{\epsilon}, j = 1, ..., n. \tag{41}$$

We may write all the equations in (41) together as $\tilde{U} = \tilde{Z}\beta + e$, where

$$\tilde{U} = \begin{bmatrix} \tilde{y}_1 - \bar{\tilde{y}} \\ \vdots \\ \tilde{y}_n - \bar{\tilde{y}} \end{bmatrix}, \ \tilde{Z} = \begin{bmatrix} \tilde{x}_{11} - \bar{\tilde{x}}_1 & ... & \tilde{x}_{k1} - \bar{\tilde{x}}_k \\ \vdots & ... & \vdots \\ \tilde{x}_{1n} - \bar{\tilde{x}}_1 & ... & \tilde{x}_{kn} - \bar{\tilde{x}}_k \end{bmatrix}, \ e = \begin{bmatrix} \tilde{e}_1 \\ \vdots \\ \tilde{e}_n \end{bmatrix}, \ \beta = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}.$$

Note that $\tilde{U}$ is an $n \times 1$ matrix, $\tilde{Z}$ is an $n \times k$ matrix, β is a $k \times 1$ matrix, and $\tilde{e}$ is an $n \times 1$ matrix. Then, the sum of squares of the absolute values of the errors is the following:

$$\tilde{e}^*\tilde{e} = (\tilde{U} - \tilde{Z}\beta)^*(\tilde{U} - \tilde{Z}\beta). \tag{42}$$

In the least square procedure, we minimize this error sum of squares of the absolute values of the errors, and then, estimate the parameter vector β. Note that $(\tilde{U} - \tilde{Z}\beta)^* = \tilde{U}^* - \beta^*\tilde{Z}^*$ and

$$\frac{\partial \tilde{e}^*\tilde{e}}{\partial \beta} = O \Rightarrow O - O - 2\tilde{Z}^*\tilde{U} + 2\tilde{Z}^*\tilde{Z}\beta = O$$

$$\beta = (\tilde{Z}^*\tilde{Z})^{-1}\tilde{Z}^*\tilde{U} \tag{43}$$

where we have assumed that $\tilde{Z}^*\tilde{Z}$ is a nonsingular matrix because $\tilde{x}_{rj}$'s are preassigned numbers, and hence, $\tilde{Z}$ can be taken as a full-rank matrix with rank $k < n$. Then, the estimated β, as per the least square estimate, again denoted by β, is $\beta = (\tilde{Z}^*\tilde{Z})^{-1}\tilde{Z}^*\tilde{U}$ and the estimated model for $\tilde{y}$ is $a_0^c + \beta^*\tilde{X}$, $\tilde{X}' = [\tilde{x}_1, ..., \tilde{x}_k]$, $\beta^* = \tilde{U}^*\tilde{Z}(\tilde{Z}^*\tilde{Z})^{-1}$ or the estimated $\tilde{y}$ is

$$\tilde{y} = a_0^c + \beta^*\tilde{X} = a_0^c + \tilde{U}^*\tilde{Z}(\tilde{Z}^*\tilde{Z})^{-1}\tilde{X} \tag{44}$$

where β is available from (43) and $a_0^c = \bar{\tilde{y}} - \beta^*\bar{\tilde{X}}$, $\bar{\tilde{X}} = [\bar{\tilde{x}}_1, ..., \bar{\tilde{x}}_k]'$. This shows that the least square procedure in the complex domain also runs parallel to that in the real domain. If $\tilde{e}$ is assumed to have an n-variate complex Gaussian distribution, then the inference problems also runs parallel to those in the real domain.

6. Concluding Remarks

In this paper, we have introduced vector/matrix differential operators in the complex domain. These differential operators in the complex domain are believed to be new. With the help of these operators, we have examined the optimization of a linear form with Hermitian-form constraint, optimization of a Hermitian form with linear form as well as Hermitian-form constraint, and optimization of a bilinear form with Hermitian-form constraints, where the linear forms and bilinear forms involve vectors and matrices in the complex domain. As applications of these optimization problems, we have extended principal component analysis and canonical correlation analysis to the complex domain. Also extended to the complex domain are the Cramer–Rao inequality, the Cauchy–Schwarz inequality, minimum-variance unbiased estimation, and least square analysis. If we use the general definition of a density $f(X)$ as a real-valued scalar function such that $f(X) \geq 0$ in the domain of X and $\int_X f(X)dX = 1$, where the argument X may be scalar or vector or matrix or a sequence of matrices in the real or complex domains [8], then the structures of the joint density, marginal density, conditional density, etc., will be parallel to those in the real domain. Then, we will be able to extend Bayesian analysis to the complex domain. One can also explore extending other multivariate statistical techniques such as factor analysis, classification problems, cluster analysis, analysis of variance, analysis of covariance, etc., to the complex domain. These are some of the open problems. Since the likelihood function L is a product of densities at the observed sample point, in the simple

random sample case, this *L* will be a real-valued scalar function. Then, one can extend the maximum likelihood method of estimation to the complex domain. For example, in the *p*-variate complex Gaussian case, in the case of a simple random sample of size *n*, we have

$$\ln L = c - np \, \ln |\tilde{\Sigma}| - \mathrm{tr}(\tilde{\Sigma}^{-1}\tilde{S}) - \sum_{j=1}^{n}(\tilde{X}_j - \tilde{\mu})^*\tilde{\Sigma}^{-1}(\tilde{X}_j - \tilde{\mu})$$

where *c* is a constant, $\tilde{S}$ is the sample sum of squares and cross-products matrix, $\tilde{\mu}$ is the population mean value, and $\tilde{\Sigma} > O$ is the Hermitian positive definite covariance matrix. We have already established results on vector and matrix differential operators in the complex domain, operating on the trace and determinant involving a Hermitian positive definite matrix. Hence, all the terms in $\frac{\partial}{\partial \tilde{\mu}}[\ln L] = O$ and $\frac{\partial}{\partial \tilde{\Sigma}}[\ln L] = O$ are defined and such equations are already solved in our discussion of our operators operating on traces and determinants. Thus, we will see that $\frac{\partial}{\partial \tilde{\mu}}[\ln L] = O$ yields the sample average as the estimate/estimator of $\tilde{\mu}$, and $\frac{\partial}{\partial \tilde{\Sigma}}[\ln L] = O$, with the estimate on $\tilde{\mu}$, yields the estimate/estimator of $\tilde{\Sigma}$ as $\frac{1}{n}\tilde{S}$. One can also examine the maximum likelihood estimation involving other scalar/vector/matrix-variate densities in the complex domain. The above are some of the open problems.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Deng, X. Texture Analysis and Physical Interpretation of Polarimetric SAR Data. Ph.D Thesis, Universitat Politechnica de Catalunya, Barcelona, Spain, 2016.
2. Du, L.; Liu, H.; Wang, P.; Feng, B.; Pan, M.; Bao, Z. Noise robust radar HRRP target recognition based on multitask factor analysis with small training data size. *IEEE Trans. Signal Process.* **2012**, *60*, 3546–3560.
3. Hellings, C.; Gogler, P.; Utschick, W. Composite real principal component analysis of complex signals. In Proceedings of the 23rd European Signal Processing Conference (EUSIPCO), Nice, France, 31 August–4 September 2015; pp. 2216–2220.
4. Horel, J.D. Complex principal component anslysis:theory and examples. *J. Clim. Appl. Meteorol.* **1984**, *23*, 21660–21673. [CrossRef]
5. Katkovnik, V.; Ponomarenko, M.; Egiazarian, K. Complex-valued image denosing based on group-wise complex-domain sparsity. *arXiv* **2017**, arXiv:1711.00362vl.
6. Liu, J.; Xu, X.; Zhang, F.; Gao, Y.; Gao, W. Modeling of spatial distribution characteristics of high proportion renewable energy based on complex principal component analysis. In Proceedings of the 2020 IEEE Sustainable Power and Energy Conference (iSPEC), Chengdu, China, 23–25 November 2020; pp. 193–198.
7. Morup, M.; Madsen, K.H.; Hansen, L.K. Shifted independent component analysis. In Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation: ICA2007, London, UK, 9–12 September 2007; pp. 89–96.
8. Mathai, A.M.; Provost, S.B.; Haubold, H.J. *Multivariate Statistical Analysis in the Real and Complex Domains*; Springer Nature: Cham, Switzerland, 2022.