Wirtinger Calculus and Complex Natural Gradient Algorithm

Rafael de Carvalho Bluhm and Charles Casimiro Cavalcante

***Abstract*— This paper presents an extension of the real tangent spaces in the framework of the Wirtinger (complex) calculus, employed to obtain a generalized expression of the natural complex gradient. An expression is proposed for the natural gradient, valid for a generalized coordinate system expressed in a complex basis, which allows for a fully complex notation. The equivalence of complex and real representations is done through the complex structure that real spaces may possess, allowing the change of the used basis. Typical examples of applications are presented, illustrating the potential of the proposed approach.**

***Keywords*— Wirtinger calculus, $\mathbb{CR}$ calculus, complex gradient, augmented, adaptive filters, complex calculus, complex differential forms.**

I. Introduction

Signal processing theory makes extensive use of complex notation nowadays. This reflects the potential of complex numbers: they can carry two pieces of information in one "package". Furthermore, they are also used to represent real valued functions with complex variables in optimization problems. This led to the augmented representation, in which an element is represented by a pair of a complex number and its complex conjugated.

From this point of view, in order to keep the computations in the complex domain, the so-called Wirtinger calculus, also named $\mathbb{CR}$ calculus, emerged. It is a natural extension to the complex domain of the usual real algebra and calculus widely used in signal processing, especially when complex signals are involved. In this notation, given a generic complex signal in vector representation $\mathbf{u}$[n], the correlation $\mathbf{R} = \mathrm{E}\{\mathbf{u}[n]\mathbf{u}^H[n]\}$ – obtained using the Hermitian and pseudo-correlation $\tilde{\mathbf{R}} = \mathrm{E}\{u[n]u^T[n]\}$ by transposition [1] – is better explored using $\mathbb{CR}$ calculus.

The present work summarizes the most important properties and rules of $\mathbb{CR}$ calculus, yielding the mathematical concepts necessary for the introduction of the proposed complex natural gradient, which has an associated complex augmented metric. The literature introduces the natural real solution in [2], which depends on the use of the metric matrix [3] as the correction factor of gradient vector. In [4] the complex gradient is presented as the anti-holomorphic derivative of a function which gradient is metric-corrected. This paper demonstrates that a complex increment, of a real valued function, depends on the holomorphic and anti-holomorphic derivatives.

It is always possible to keep all calculations in real notation, treating complex values as two-dimensional real values and then use a real metric, as in [5], but it makes the mathematical representation less natural and more convoluted than the calculations in the complex representation. This allows the use of the Wirtinger calculus, as long as the complex augmented metric is known, for any given coordinate system (in even dimension), keeping the correspondence between the real case and the complex augmented case.

Lastly, to show the equivalence between the representations, three examples are introduced: in Cartesian, in cylindrical and in oblique coordinates. In the first, the presence of a scale factor makes it possible to compare, for example, the gradient descent algorithm in real and complex domains along the real axis (both are equal when the real signals are applied) which generally does not occur in several studies, such as [4] and [6] but has the correct application on [1], [7] and [8]. The last two examples demonstrate the relationship between the augmented complex notation and the natural gradient results in [2].

Notation: scalars are represented as regular lower case letters and differential forms are written as lower case Greek letters. Vectors written in capital letters and local vector coordinates with bold lower case correspondent. Matrices are written in bold upper case. The complex conjugation is represented by a bar over the scalar, vector or matrix as $\bar{()}$, the dual value by $()^*$ and the imaginary unit as j. Isomorphism is denoted by $\approx$. Transpositions are denoted by $()^T$ and the Hermitian as the transpose conjugated as $()^H$.

II. Tangent spaces, vectors and co-vectors

A. Complex Calculus

CR calculus comes from the well established theory of complex analysis, where a complex number is an element $z = x + jy \in \mathbb{C}$, for $x, y \in \mathbb{R}$, that can be extended to vector complex spaces by taking the coordinate transformation $(x, y) \to (z, \bar{z})$. An important result in differential form, writing $dz = dx + jdy$ and $d\bar{z} = dx - jdy$, yields the exterior derivative of a function (complex or real)

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial \bar{z}}d\bar{z} \ , \tag{1}$$

defining $\partial/\partial z = 1/2(\partial/\partial x - j\partial/\partial y)$ and its conjugated $\partial/\partial\bar{z} = 1/2(\partial/\partial x + j\partial/\partial y)$. Note that the coordinates $(z, \bar{z})$ are dependent, but the forms dz and $d\bar{z}$ are linearly independent. This means that differential calculus can be done with this choice of complex coordinates. The pair $(z, \bar{z})$ has a restriction in the second variable, and for this reason, the space is also denoted as $(z, \bar{z}) \in \mathbb{C}^2_* \approx \mathbb{C} \approx \mathbb{R}^2$.

Using this notation, the Cauchy-Riemann condition is written as

$$\frac{\partial f}{\partial \bar{z}} = 0 \ , \tag{2}$$

for a holomorphic function, whose derivative is given by $f'(z) = \partial f/\partial z$. The problem is that most cost functions (nonzero or nonconstant ones) are not holomorphic, not accepting the usual complex derivative. Due to this fact a new approach for the derivative, based on the complex differential forms, is needed.

Rafael de Carvalho Bluhm and Charles Casimiro Cavalcante, Federal University of Ceará (UFC), Fortaleza-CE, Brazil, E-mail: faelbluhm@gmail.com,charles@gtel.ufc.br.

B. Change of Basis

The previous definitions allow to build a complex vector space from a real space. This becomes clear when considering the vector basis of the tangent plane TM^{2n} [9] at some valid point (fixed, but not specified) of a manifold M^{2n},

$$\left\{ \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}, \frac{\partial}{\partial y_1} \ldots, \frac{\partial}{\partial y_n} \right\} ,$$

and rewriting it with the help of the differential complex operators as in [10],

$$\left\{ \frac{\partial}{\partial z_k}, \frac{\partial}{\partial \bar{z}_k} \right\} = \left\{ \frac{1}{2}\left(\frac{\partial}{\partial x_k} - j\frac{\partial}{\partial y_k} \right), \frac{1}{2}\left(\frac{\partial}{\partial x_k} + j\frac{\partial}{\partial y_k} \right) \right\} . \tag{3}$$

These complex basis are dual to the dual bases $\{dz_1, \ldots, dz_n, d\bar{z}_1, \ldots, d\bar{z}_n\}$, where $dz_k = dx_k + jdy_k$ and $d\bar{z}_k = dx_k - jdy_k$ for $k = \{1, \ldots, n\}$. This can be verified by direct substitution, i.e

$$dz_m\left(\frac{\partial}{\partial z_n}\right) = \delta_{mn} \ , \quad dz_m\left(\frac{\partial}{\partial \bar{z}_n}\right) = 0$$

$$d\bar{z}_m\left(\frac{\partial}{\partial z_n}\right) = 0 \ , \quad d\bar{z}_m\left(\frac{\partial}{\partial \bar{z}_n}\right) = \delta_{mn} \ ,$$

where $\delta_{mn} = 1$ if $m = n$, and zero otherwise. It is possible set three different constructions using the complex basis: the real representation with complex numbers, complex coefficients with real basis and the fully complex – coefficients and basis – the last two equivalent.

C. Complexified vector spaces

The representation of complex coefficients over real basis at a given point over a $2n$-dimensional manifold gives us the called complexified tangent space. For the complex coefficients $a_k = a_k^r + ja_k^i$, $b_k = b_k^r + jb_k^i \in \mathbb{C}$, a tangent vector U it is written as

$$U = \sum_{k=1}^{n} \left(a_k \frac{\partial}{\partial x_k} + b_k \frac{\partial}{\partial y_k} \right) \ .$$

Changing basis according (3), the same vector is now written as

$$U = \sum_{k=1}^{n} \left(\lambda_k \frac{\partial}{\partial z_k} + \mu_k \frac{\partial}{\partial \bar{z}_k} \right) \ ,$$

with the coefficients

$$\begin{cases} \lambda_k = a_k^r - b_k^i + j(a_k^i + b_k^r) \\ \mu_k = a_k^r + b_k^i + j(a_k^i - b_k^r) \ , \end{cases} \tag{4}$$

representing the complex coefficients over the complex basis. This tangent vector in local coordinates is represented as a column vector as

$$\mathbf{u} = \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} \ . \tag{5}$$

where $\boldsymbol{\lambda} = [\lambda_1 \ldots \lambda_n]^T$ and $\boldsymbol{\mu} = [\mu_1 \ldots \mu_n]^T$. This vector belongs to $T_{\mathbb{C}}M = TM \otimes \mathbb{C}$, isomorphic to $\mathbb{C}^{2n} \approx \mathbb{R}^{4n}$. In this notation the vectors are called here *complexified.*

D. Real vectors representation in augmented notation

For a real tangent vector $V \in TM \approx \mathbb{R}^{2n}$, its representation in real tangent basis is

$$V = \sum_{k=1}^{n} \left(r_k \frac{\partial}{\partial x_k} + s_k \frac{\partial}{\partial y_k} \right) \ ,$$

for $r_k,\ s_k \in \mathbb{R}$. This same vector in complex base is written as

$$V = \sum_{k=1}^{n} \left[(r_k + js_k)\frac{\partial}{\partial z_k} + (r_k - js_k)\frac{\partial}{\partial \bar{z}_k} \right] \ .$$

Taking $\eta_k = r_k + js_k$, the complex representation of a real vector V is an augmented complex vector,

$$V = \sum_{k=1}^{n} \left(\eta_k \frac{\partial}{\partial z_k} + \bar{\eta}_k \frac{\partial}{\partial \bar{z}_k} \right) \ , \tag{6}$$

which can be represented in local coordinates as

$$\mathbf{v} = \begin{bmatrix} \boldsymbol{\eta} \\ \bar{\boldsymbol{\eta}} \end{bmatrix} \ , \tag{7}$$

where $\boldsymbol{\eta} = [\eta_1 \ldots \eta_n]^T$. Note that, it is possible to write a vector only with the holomorphic basis elements $\partial/\partial z_k$ as

$$U' = \sum_{k=1}^{n} \lambda_k \frac{\partial}{\partial z_k} \ ,$$

calling it as a holomorphic tangent space $T_{\mathbb{C}}^{1,0}M$. Doing the same for anti-holomorphic part

$$U'' = \sum_{k=1}^{n} \mu_k \frac{\partial}{\partial \bar{z}_k} \ ,$$

results in what can be called as $T_{\mathbb{C}}^{0,1}M$ the anti-holomorphic space for the base elements $\partial/\partial \bar{z}_k$. With these two it is possible to write the relation $T_{\mathbb{C}}M = T_{\mathbb{C}}^{1,0}M \oplus T_{\mathbb{C}}^{0,1}M$.

When expanding basis vectors and coefficients it is important take note about the complex structure J as an operator that take the bases vectors into its pairs

$$\begin{cases} J\frac{\partial}{\partial x_k} = \frac{\partial}{\partial y_k} \\ J\frac{\partial}{\partial y_k} = -\frac{\partial}{\partial x_k} \ . \end{cases} \tag{8}$$

Considering the general mnemonic rule for the product $1 \cdot j = j \cdot 1 = J$ and $j \cdot j = -1$, it is possible to represent real vectors only by the holomorphic component

$$V' = \sum_{k=1}^{n} \eta_k \frac{\partial}{\partial z_k} = \sum_{k=1}^{n} \frac{1}{2}(1r_k + js_k) \cdot \left(1\frac{\partial}{\partial x_k} - j\frac{\partial}{\partial y_k} \right)$$

$$V' = \sum_{k=1}^{n} \frac{1}{2} \left(r_k \frac{\partial}{\partial x_k} - r_k J\frac{\partial}{\partial y_k} + s_k J\frac{\partial}{\partial x_k} + s_k \frac{\partial}{\partial y_k} \right) \ .$$

After using the relations at (8) (8), results into

$$V' = \sum_{k=1}^{n} \left(r_k \frac{\partial}{\partial x_k} + s_k \frac{\partial}{\partial y_k} \right) = V \ .$$

Vectors in (6) and (7) formats are called *augmented* vectors. The different notations used to express a vector with the basis and in column (matrix) have a reason: while for augmented representation the condition $\bar{V} = V$ holds (basis possess complex conjugate pairs), in column notation, in general, $\bar{\mathbf{v}} \neq \mathbf{v}$.

E. The representations for 1-forms

1-forms are vectors in the dual space of tangent spaces. When they act on tangent vectors, these forms result into a scalar value. A general 1-form with complex coefficients $a_k^* = a_k^{*r} + ja_k^{*i}$ and $b_k^* = b_k^{*r} + jb_k^{*i}$ both in $\mathbb{C}$, can be expressed as

$$\upsilon = \sum_{k=1}^{n} \left(a_k^* dx_k + b_k^* dy_k\right) \ .$$

With $dz_k = dx_k + jdy_k$ and its complex conjugated the same form can be written as

$$\upsilon = \frac{1}{2} \sum_{k=1}^{n} \left(\mu_k^* dz_k + \lambda_k^* d\bar{z}_k\right) \ ,$$

for coefficients

$$\begin{cases} \lambda_k^* = a_k^{*r} - b_k^{*i} + j(a_k^{*i} + b_k^{*r}) \\ \mu_k^* = a_k^{*r} + b_k^{*i} + j(a_k^{*i} - b_k^{*r}) \ . \end{cases} \tag{9}$$

To hold the Hermitian property while working with the complex augmented notation, one can note that the application of a 1-form requires its conjugated (including basis). For a vector $Q = \Sigma_{k=1}^{n}(\delta_k \partial/\partial z_k + \epsilon_k \partial/\partial \bar{z}_k)$ this application is

$$\bar{\upsilon}(Q) = \frac{1}{2} \sum_{k=1}^{n} \left(\bar{\lambda^*}_k \delta_k + \bar{\mu^*}_k \epsilon_k\right) = \frac{1}{2} \begin{bmatrix} \bar{\boldsymbol{\lambda}^*} & \bar{\boldsymbol{\mu}^*} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\epsilon} \end{bmatrix} \tag{10}$$

for row vectors $\boldsymbol{\lambda}^* = [\lambda_1^* \cdots \lambda_n^*]$ and $\boldsymbol{\mu}^* = [\mu_1^* \cdots \mu_n^*]$. This expression in local coordinates is a row vector given by

$$\boldsymbol{\upsilon} = \frac{1}{2} \mathbf{u}^* \ ,$$

for the row vector $\mathbf{u}^* = [\boldsymbol{\lambda}^* \ \ \boldsymbol{\mu}^*]$. This will be important soon due the relation between $\mathbf{u}^*$ and the vector (5) through the metric.

For the complex augmented case just take the coordinates as $\bar{\omega} = [\boldsymbol{\alpha}^* \ \ \bar{\boldsymbol{\alpha}^*}]$ and verify that

$$\bar{\omega}(V) = \frac{1}{2} \bar{\mathbf{w}^*} \mathbf{v} \tag{11}$$

as the associated complex augmented row form.

The application of a real differential 1-form defined as $\omega = \Sigma_{k=1}^{n}(p_k^* dx_k + q_k^* dy_k)$ in a vector $V = \Sigma_{k=1}^{n}(r_k \partial/\partial x_k + s_k \partial/\partial y_k)$ results into a real value

$$\bar{\omega}(V) = \begin{bmatrix} \mathbf{p}^* & \mathbf{q}^* \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} = \mathbf{p}^* \mathbf{r} + \mathbf{q}^* \mathbf{s} \ , \tag{12}$$

also valid for the complex augmented representation, with $\alpha_k^* = p_k^* + jq_k^*$ and $\eta_k = r_k + js_k$,

$$\bar{\omega}(V) = \frac{1}{2} \begin{bmatrix} \bar{\boldsymbol{\alpha}^*} & \boldsymbol{\alpha}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \bar{\boldsymbol{\eta}} \end{bmatrix} = \Re\left\{\bar{\boldsymbol{\alpha}^*} \boldsymbol{\eta}\right\} = \mathbf{p}^* \mathbf{r} + \mathbf{q}^* \mathbf{s} \ . \tag{13}$$

The results in (12) and in (13) shows the equivalence between representations: in complex augmented representation, the action of a 1-form on a tangent augmented vector results into a real number as in real case. Note that the choice of the 1-form complex conjugate acting on vectors implies the use of an inner product that can be sesquilinear on a second operand: allowing to use complex matrix notations in complex augmented notation.

III. Co-gradient and the gradient

A. Inner Product and Metric

The sesquilinear form h applied to the vectors $V = \sum(\eta_k \partial/\partial z_k + \bar{\eta}_k \partial/\partial \bar{z}_k)$ and $W = \sum(\alpha_k \partial/\partial z_k + \bar{\alpha}_k \partial/\partial \bar{z}_k)$ for $\eta_k, \alpha_k \in \mathbb{C}$ allow us to write

$$h(V, W) = \sum_{i,k} \left\langle \eta_i \frac{\partial}{\partial z_i} + \bar{\eta}_i \frac{\partial}{\partial \bar{z}_i}, \alpha_k \frac{\partial}{\partial z_k} + \bar{\alpha}_k \frac{\partial}{\partial \bar{z}_k} \right\rangle \ ,$$

in matrix notation as

$$h(V, W) = \begin{bmatrix} \boldsymbol{\alpha}^H & \boldsymbol{\alpha}^T \end{bmatrix} \cdot \tilde{\mathbf{H}} \cdot \begin{bmatrix} \boldsymbol{\eta} \\ \bar{\boldsymbol{\eta}} \end{bmatrix} = \mathbf{w}^H \tilde{\mathbf{H}} \mathbf{v} \ ,$$

for the compose matrix defined as

$$\tilde{\mathbf{H}} = \begin{bmatrix} \left[\left\langle \frac{\partial}{\partial z_i}, \frac{\partial}{\partial z_k} \right\rangle\right] & \left[\left\langle \frac{\partial}{\partial \bar{z}_i}, \frac{\partial}{\partial z_k} \right\rangle\right] \\ \left[\left\langle \frac{\partial}{\partial z_i}, \frac{\partial}{\partial \bar{z}_k} \right\rangle\right] & \left[\left\langle \frac{\partial}{\partial \bar{z}_i}, \frac{\partial}{\partial \bar{z}_k} \right\rangle\right] \end{bmatrix} \ , \tag{14}$$

where the index i represents the column and k the line. Note that $\tilde{\mathbf{H}} = \tilde{\mathbf{H}}^H$, a Hermitian symmetric matrix.

Given a real metric

$$\mathbf{G} = \begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}^T & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \left[\left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_k} \right\rangle\right] & \left[\left\langle \frac{\partial}{\partial y_i}, \frac{\partial}{\partial x_k} \right\rangle\right] \\ \left[\left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial y_k} \right\rangle\right] & \left[\left\langle \frac{\partial}{\partial y_i}, \frac{\partial}{\partial y_k} \right\rangle\right] \end{bmatrix}$$

for $\mathbf{P} = \mathbf{P}^T$, $\mathbf{R} = \mathbf{R}^T$ and $\mathbf{G} = \mathbf{G}^T$, it can be verified, by direct expansion of (14) and for $\mathbf{H} = 2\tilde{\mathbf{H}}$, that

$$\mathbf{H} = \frac{1}{2} \begin{bmatrix} \mathbf{P} + \mathbf{R} + j(\mathbf{Q}^T - \mathbf{Q}) & \mathbf{P} - \mathbf{R} + j(\mathbf{Q}^T + \mathbf{Q}) \\ \mathbf{P} - \mathbf{R} - j(\mathbf{Q}^T + \mathbf{Q}) & \mathbf{P} + \mathbf{R} - j(\mathbf{Q}^T - \mathbf{Q}) \end{bmatrix} . \tag{15}$$

The last expression shows the relation between complex and real form for the metric matrix, also used to ensure the differential form

$$h(V, W) = \langle V, W \rangle = \frac{1}{2} \mathbf{w}^H \mathbf{H} \mathbf{v} \ , \tag{16}$$

is valid to the real coefficients representation. The metric is related to the transformation between vectors and their dual. Comparing (16) with the expression (11) is easy to see that the application of a form can be represented as an Hermitian product

$$\bar{\omega}(V) = \langle V, W \rangle \ .$$

It can be noted that $\mathbf{H}$ is the metric that preserves the real Euclidean distance while working in augmented notation as in (12) and (13).

B. The Co-Gradient

The differential form of a complex valued function is a 1-form, called exterior derivative. In complex coordinates it is expressed in a compact notation as

$$df = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}} & \frac{\partial f}{\partial \bar{\mathbf{z}}} \end{bmatrix} \begin{bmatrix} d\mathbf{z} \\ d\bar{\mathbf{z}} \end{bmatrix} \ . \tag{17}$$

for the row vectors $\partial f/\partial \mathbf{z} = [\partial f/\partial z_1 \dots \partial f/\partial z_n]$, $\partial f/\partial \bar{\mathbf{z}} = [\partial f/\partial \bar{z}_1 \dots \partial f/\partial \bar{z}_n]$ and for $d\mathbf{z} = [dz_1 \dots dz_n]^T$ and its complex conjugated. This differential form operates on a vector V, resulting in the expression

$$df(V) = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}} & \frac{\partial f}{\partial \bar{\mathbf{z}}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \bar{\boldsymbol{\eta}} \end{bmatrix} \tag{18}$$

in local coordinates. For a real function f, the following is valid

$$\overline{\left(\frac{\partial f}{\partial \mathbf{z}}\right)} = \frac{\partial \bar{f}}{\partial \bar{\mathbf{z}}} = \frac{\partial f}{\partial \bar{\mathbf{z}}} \ . \tag{19}$$

In this case, the differential is an augmented form, and it can be written as

$$df(V) = \frac{\partial f}{\partial \mathbf{z}} \boldsymbol{\eta} + \left(\frac{\partial f}{\partial \mathbf{z}} \boldsymbol{\eta}\right)^* = 2\mathrm{Re}\left\{\frac{\partial f}{\partial \mathbf{z}} \boldsymbol{\eta}\right\} \ ,$$

which is not valid for a general complexified vector. For a differential form as expressed in (17), in general $\overline{(df)} \neq df$. The equality only holds for real functions. For this reason, and also to keep the isomorphism to the $\mathbb{R}^{2n}$, from now on in this work the complex augmented configuration will be considered for vectors, forms and all functions as real valued.

C. The Gradient

Property (16) allows the representation of the gradient ∇f in a vector version of the differential form df. For a real function f and an augmented vector V, directional derivative is

$$df(V) = \langle V, \nabla f \rangle = \frac{1}{2} (\boldsymbol{\nabla} f)^H \mathbf{H} \cdot \mathbf{v}. \tag{20}$$

The dual vector df is a 1-form, a row vector, already scaled and complex conjugated by definition. Naming the augmented co-gradient row operator as $\boldsymbol{\delta} f = [\partial f/\partial \mathbf{z} \ \ \partial f/\partial \bar{\mathbf{z}}]$ one can write now

$$\boldsymbol{\delta} f \cdot \mathbf{v} = \frac{1}{2} (\boldsymbol{\nabla} f)^H \mathbf{H} \cdot \mathbf{v} \ ,$$

valid for any vector input. This leads to the proposed complex augmented gradient, in short notation

$$\boldsymbol{\nabla} f = 2\mathbf{H}^{-1} (\boldsymbol{\delta} f)^H \ , \tag{21}$$

that depends on both holomorphic and anti-holomorphic derivatives. Expanding (21) into complex components

$$\boldsymbol{\nabla} f = \begin{bmatrix} \boldsymbol{\nabla} f_{\mathbf{z}} \\ \boldsymbol{\nabla} f_{\bar{\mathbf{z}}} \end{bmatrix} = 2\mathbf{H}^{-1} \begin{bmatrix} \left(\frac{\partial f}{\partial \mathbf{z}}\right)^H \\ \left(\frac{\partial f}{\partial \bar{\mathbf{z}}}\right)^H \end{bmatrix} = 2\mathbf{H}^{-1} \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}^H} \\ \frac{\partial}{\partial \mathbf{z}^T} \end{bmatrix} f \ , \tag{22}$$

where the indexes $\mathbf{z}$ and $\bar{\mathbf{z}}$ refer to the corresponding basis components. Note that the relation $\overline{\boldsymbol{\nabla} f_{\mathbf{z}}} = \boldsymbol{\nabla} f_{\bar{\mathbf{z}}}$ is valid for real functions, due the format of (15) whose inverse keeps the augmented format [11]. For an inverse metric matrix represented by

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \bar{\mathbf{N}} & \bar{\mathbf{M}} \end{bmatrix} \ ,$$

the first component of gradient is the complex natural gradient expressed as

$$\boldsymbol{\nabla} f_{\mathbf{z}} = 2\mathbf{M} \cdot \frac{\partial f}{\partial \mathbf{z}^H} + 2\mathbf{N} \cdot \frac{\partial f}{\partial \mathbf{z}^T} \ . \tag{23}$$

Here we have two equivalent definitions: the proposed complex natural augmented gradient that is given by expression (21) and its first component: the complex natural gradient expressed in (23). To clarify the concept of this gradient, basic examples of the its use are presented in the next section.

IV. Realizations of the Complex Gradient and the Application to Gradient Descent

A. Cartesian coordinates representation

For for a metric $\mathbf{G} = \mathbf{I}$ (the identity matrix) in $\mathbb{R}^{2N}$, the complex natural gradient (the component of non-conjugated base) is

$$\boldsymbol{\nabla} f_{\mathbf{z}} = 2 \frac{\partial f}{\partial \mathbf{z}^H}$$

which yields the Cartesian complex natural gradient as in [2], but in the complex framework. As an application of this gradient, consider the steepest descent algorithm

$$\mathbf{z}_{k+1} = \mathbf{z}_k - \mu_k 2 \frac{\partial f}{\partial \mathbf{z}^H} \ , \tag{24}$$

for a real scalar factor μ_k. It can be verified the equivalence to the real $\mathbb{R}^{2N}$ case by taking $\mathbf{z} = \mathbf{x} + j\mathbf{y} \in \mathbb{C}^N$

$$\mathbf{x}_{k+1} + j\mathbf{y}_{k+1} = \mathbf{x}_k + j\mathbf{y}_k + \mu_k 2 \left[\frac{1}{2}\left(\frac{\partial J}{\partial \mathbf{x}^T} + j \frac{\partial J}{\partial \mathbf{y}^T}\right)\right]$$

resulting into $\mathbf{x}_{k+1} = \mathbf{x}_k + \mu_k \boldsymbol{\nabla}_{\mathbf{x}} J$ and $\mathbf{y}_{k+1} = \mathbf{y}_k + \mu_k \boldsymbol{\nabla}_{\mathbf{y}} J$, for $\boldsymbol{\nabla}_{\mathbf{x}} = \partial/\partial \mathbf{x}^T$ and $\boldsymbol{\nabla}_{\mathbf{y}} = \partial/\partial \mathbf{y}^T$ due to the identity metric on Cartesian coordinates.

B. Polar coordinates representation

We can consider another example taking the polar two-dimensional coordinates where $\mathbf{G} = \mathrm{diag}\{1, \rho^2\}$, for a coordinate pair $(\rho \cos\varphi, \rho \sin\varphi)$. By direct substitution in (15) and then into (21)

$$\begin{bmatrix} \nabla f_z \\ \nabla f_{\bar{z}} \end{bmatrix} = 2\mathbf{H}^{-1} \begin{bmatrix} \frac{\partial}{\partial \bar{z}} \\ \frac{\partial}{\partial z} \end{bmatrix} = 2\left(\frac{1}{2}\begin{bmatrix} 1 + \frac{1}{\rho^2} & 1 - \frac{1}{\rho^2} \\ 1 - \frac{1}{\rho^2} & 1 + \frac{1}{\rho^2} \end{bmatrix}\right) \begin{bmatrix} \frac{\partial}{\partial \bar{z}} \\ \frac{\partial}{\partial z} \end{bmatrix} \ .$$

The first component of the augmented gradient in complex notation is

$$\nabla f_z = \left(1 + \frac{1}{\rho^2}\right) \frac{\partial f}{\partial \bar{z}} + \left(1 - \frac{1}{\rho^2}\right) \frac{\partial f}{\partial z} \ ,$$

with the presence of $\partial f/\partial z$ and $\partial f/\partial \bar{z}$. This can be verified by direct substitution of basis, resulting in the expression of the polar non-normalized basis

$$\nabla f_z = \frac{\partial f}{\partial \rho} + j \frac{1}{\rho^2} \frac{\partial f}{\partial \varphi} \ .$$

Writing the gradient in the associated complex basis and then using the complex structure (8)

$$\nabla f_z \frac{\partial}{\partial z} = \left(\frac{\partial f}{\partial \rho} + j\frac{1}{\rho^2}\frac{\partial f}{\partial \varphi}\right) \cdot \frac{1}{2}\left(\frac{\partial}{\partial \rho} - j\frac{\partial}{\partial \varphi}\right)$$

$$= \frac{1}{2}\left(\frac{\partial f}{\partial \rho}\frac{\partial}{\partial \rho} - \frac{\partial f}{\partial \rho}J\frac{\partial}{\partial \varphi} + \frac{1}{\rho^2}\frac{\partial f}{\partial \varphi}J\frac{\partial}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial f}{\partial \varphi}\frac{\partial}{\partial \varphi}\right)$$

$$\nabla f_z \frac{\partial}{\partial z} = \frac{\partial f}{\partial \rho}\frac{\partial}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial f}{\partial \varphi}\frac{\partial}{\partial \varphi} \ .$$

The presence of the complex structure allows writing the non-normalized basis $\{\partial/\partial\rho, \partial/\partial\varphi\}$ locally as the complex base $\{1, j\}$. Note that some vector V is written directly into the given coordinates as $V = V_\rho + jV_\varphi$. As an application of this gradient update, consider the following adjusting rule for small displacements

$$\delta z = -\mu \nabla f_z$$

that provides the expanded version

$$\delta\rho + j\delta\varphi = -\mu\left(\frac{\partial f}{\partial \rho} + j\frac{1}{\rho^2}\frac{\partial f}{\partial \varphi}\right) \ ,$$

of a given $z_{k+1} = z_k - \mu_k \nabla f_z$, as in [2] at a instant k. Note that the natural gradient may be different if we use $z = x(\rho,\varphi) + jy(\rho,\varphi) = \rho\cos\varphi + j\rho\sin\varphi$ as Cartesian to polar coordinates. In this case

$$\nabla_z = \frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y} \ ,$$

related to the Jacobian transformation

$$\begin{bmatrix}\nabla_z \\ \nabla_{\bar z}\end{bmatrix} = \begin{bmatrix}\cos\varphi + j\sin\varphi & -\rho\sin\varphi + j\cos\varphi \\ \cos\varphi - j\sin\varphi & -\rho\sin\varphi - j\rho\cos\varphi\end{bmatrix}\begin{bmatrix}\frac{\partial f}{\partial \rho} \\ \frac{1}{\rho^2}\frac{\partial f}{\partial \varphi}\end{bmatrix} \ ,$$

shortened as

$$\begin{bmatrix}\nabla_z \\ \nabla_{\bar z}\end{bmatrix} = \frac{\partial(z,\bar z)}{\partial(x,y)} \cdot \frac{\partial(x,y)}{\partial(\rho,\varphi)} \cdot \nabla_{(\rho,\varphi)} f = \frac{\partial(z,\bar z)}{\partial(\rho,\varphi)} \cdot \nabla_{(\rho,\varphi)} f \ .$$

C. Oblique representation

All previous gradient examples were presented using orthogonal coordinate systems ($\mathbf{Q} = 0$ on (15)). This example deals with a simple two-dimensional oblique system where $Q \neq 0$, a coordinate system $(u + v\cos\varphi, v\sin\varphi)$ for a fixed value of φ. From the real metric matrix

$$\mathbf{G} = \begin{bmatrix}1 & \cos\varphi \\ cos\varphi & 1\end{bmatrix} \ ,$$

we can compute $\mathbf{H}$ using (15) and then the complex augmented gradient with (22)

$$\begin{bmatrix}\nabla f_{\mathbf{z}} \\ \nabla f_{\bar{\mathbf{z}}}\end{bmatrix} = 2 \cdot \left(\frac{1}{2}\begin{bmatrix}2 & 2j\cos\varphi \\ -2j\cos\varphi & 2\end{bmatrix}\right)^{-1}\begin{bmatrix}\frac{\partial f}{\partial \bar z} \\ \frac{\partial f}{\partial z}\end{bmatrix} \ .$$

The result of the first line provides the complex gradient:

$$\nabla f_z = \frac{2}{\sin^2\varphi}\left(\frac{\partial f}{\partial \bar z} - j\cos\varphi\frac{\partial f}{\partial z}\right) \ .$$

This result agrees with the real case through the use of the complex structure

$$\nabla f_z = \frac{1}{\sin^2\varphi}\left[\frac{\partial f}{\partial u} - \cos\varphi\frac{\partial f}{\partial v} + j\left(-\cos\varphi\frac{\partial f}{\partial u} + \frac{\partial f}{\partial v}\right)\right] \ ,$$

representing the gradient of the oblique representation: the real part is the update on u component and the imaginary for v, for small displacements $\delta z = \delta u + i\delta v$.

V. Conclusion

This work applied the Wirtinger calculus of complex augmented vectors to propose the complex natural gradient (and the augmented version), which is related to inverse of the augmented metric. The natural gradient is corrected by this inverse and pondered by its holomorphic and anti-holomorphic derivatives. This was made possible by studying differential forms and defining a Hermitian inner product over augmented vectors.

Even in the case where the Cartesian metric is the identity, the complex has an associated scale factor. This factor does not have any effect on a minimum (or maximum) real valued problem like $\partial f/\partial z^* = 0$, but has effect on the update rate of the gradient. At [6] there is some confusion about the factor of 2 associated at some choice, such as "$\partial z/\partial z = 2$", while the factor of 2 comes from the metric scaling factor $\tilde{\mathbf{H}}$ (for Cartesian $\mathbf{G}$) as this work shows. The relation $\partial z/\partial z = 1$ is always true and can be verified with (3).

For the polar and oblique cases, it was seen that the gradient depends on holomorphic and anti-holomorphic derivatives, even though it is an orthogonal system. Hence, the augmenrted complex gradient, which preserves distances for any coordinate system, is adequate for the general complex cases.

Acknowledgments

The authors would like to thank Mr. Thiago Azevedo de Vasconcelos by his insights and suggestion which improved the clarity of the text. They also thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and CNPq (Proc. 313151/2020-2 and 409228/2021-4) for the partial financial support.

References

[1] Mandic, Danilo and Goh, Vanessa (Su Lee) *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models (Adaptive and Learning Systems for Signal Processing, Communications and Control Series)*. Wiley 2009.

[2] Amari, S. and Douglas S.C. *Why Natrual Gradient?*, Acoustics, Speech, and Signal Processing, 1988. ICASSP-88. International Conference, June 1998

[3] Kühnel, Wolfgang. *Differential geometry: curves - surfaces - manifolds*, American Mathematical Society, 2005.

[4] Kreutz–Delgado, Ken *The Complex Gradient Operator and the CR-Calculus*, arXiv:0906.4835v1, 26 Jun 2009.

[5] Nitta, Tohru *Natural Gradient Descent for Training Stochastic Complex-Valued Neural Networks*, International Journal of Advanced Computer Science and Applications, Vol. 5, No. 7, 2014.

[6] D.H. Brandwood, *A complex gradient operator and its application in adaptive array theory*, IEEPROC, Vol. 130, Pts. Fand H, No. 1, February 1983.

[7] Monzingo, R.A and Miller, T.W.: *Introduction to adaptive arrays* Wiley-Interscience, 1980.

[8] Widrow, B.; McCool, J.; and Ball, M. *The complex LMS algorithm* . IEEE 63, pp. 719-720, 1975.

[9] Manfredo P. Do Carmo, *Differential Forms and Applications*, Springer-Verlag, New York, 2000.

[10] Huybrechts, Daniel *Complex Geometry: An Introduction*, Springer, 2004.

[11] Schreier, Peter J and Scharf, Louis L.*Statistical Signal Processing of Complex-Valued Data*, Cambridge University Press, 2010