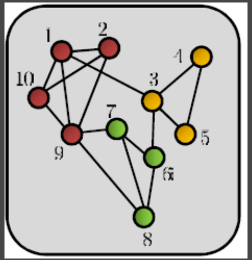

INFERENCE OVER NETWORKS

LECTURE #2: Complex Gradients Vectors

Professor Ali H. Sayed
UCLA Electrical Engineering

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Part I:
Background Material

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Reference

Appendix A (Complex Gradient Vectors, pp. 712-719):

A. H. Sayed, ``Adaptation, learning, and optimization over networks," ***Foundations and Trends in Machine Learning***, vol. 7, issue 4-5, pp. 311-801, NOW Publishers, 2014.

Setting

Let $g(z)$ denote a scalar real or complex-valued function of a complex variable, z. The function $g(z)$ need not be holomorphic in the variable z and, therefore, it need not be differentiable in the traditional complex differentiation sense (cf. definition (A.3) further ahead). In many instances though, we are only interested in determining the locations of the stationary points of $g(z)$. For these cases, it is sufficient to rely on a different notion of differentiation, which we proceed to motivate.

Cauchy-Riemann Conditions

To motivate the alternative differentiation concept, we first review briefly the traditional definition of complex differentiation. Thus, assume z is a scalar and let us express it in terms of its real and imaginary parts, denoted by x and y, respectively:

$$z \stackrel{\Delta}{=} x + jy, \quad j \stackrel{\Delta}{=} \sqrt{-1} \tag{A.1}$$

We can then interpret $g(z)$ as a two-dimensional function of the real variables $\{x, y\}$ and represent its real and imaginary parts as functions of these same variables, say, as $u(x, y)$ and $v(x, y)$:

$$g(z) \stackrel{\Delta}{=} u(x, y) \; + \; jv(x, y) \tag{A.2}$$

Cauchy-Riemann Conditions

We denote the traditional complex derivative of $g(z)$ with respect to z by $g'(z)$ and define it as the limit:

$$g'(z) \;\stackrel{\Delta}{=}\; \lim_{\Delta z \rightarrow 0} \frac{g(z+\Delta z) \;-\; g(z)}{\Delta z} \tag{A.3}$$

or, more explicitly,

$$g'(z) \;=\; \lim_{\Delta z \rightarrow 0} \frac{g(x+\Delta x, y+\Delta y) \;-\; g(x,y)}{\Delta x + j\Delta y} \tag{A.4}$$

Cauchy-Riemann Conditions

where we are writing $\Delta z = \Delta x + j\Delta y$. For $g(z)$ to be differentiable at location z, in which case it is also said to be *holomorphic* at z, then the above limit needs to exist regardless of the direction from which $z + \Delta z$ approaches z. In particular, if we set $\Delta y = 0$ and let $\Delta x \to 0$, then the above definition gives that $g'(z)$ should be equal to

$$g'(z) \;=\; \frac{\partial u(x,y)}{\partial x} \;+\; j\frac{\partial v(x,y)}{\partial x} \tag{A.5}$$

Cauchy-Riemann Conditions

On the other hand, if we set $\Delta x = 0$ and let $\Delta y \to 0$ so that $\Delta z = j\Delta y$, then the definition gives that the same $g'(z)$ should be equal to

$$g'(z) \;=\; \frac{\partial v(x,y)}{\partial y} \;-\; j\frac{\partial u(x,y)}{\partial y} \qquad \text{(A.6)}$$

Expressions (A.5) and (A.6) must coincide, which means that the real and imaginary parts of $g(z)$ should satisfy the conditions:

Cauchy-Riemann Conditions

$$\begin{cases} \dfrac{\partial u(x,y)}{\partial x} = \dfrac{\partial v(x,y)}{\partial y} \\[2ex] \dfrac{\partial u(x,y)}{\partial y} = -\dfrac{\partial v(x,y)}{\partial x} \end{cases} \tag{A.7}$$

These are known as the *Cauchy-Riemann* conditions [5, 198]. It can be shown that these conditions are not only necessary for a complex function $g(z)$ to be differentiable at location z, but if the partial derivatives of $u(x,y)$ and $v(x,y)$ are continuous, then they are also sufficient.

Example #A.1

Example A.1 (Real-valued functions). Consider the quadratic function $g(z) = |z|^2$. It is straightforward to verify that $g(x,y) = x^2 + y^2$ so that

$$u(x,y) = x^2 + y^2, \qquad v(x,y) = 0 \tag{A.8}$$

Therefore, the Cauchy-Riemann conditions (A.7) are not satisfied in this case (except at the point $x = y = 0$). More generally, it is straightforward to verify that any other (nonconstant) real-valued function, $g(z)$, cannot satisfy (A.7) except possibly at some locations. It turns out though that real-valued cost functions of this form are commonplace in problems involving estimation,

Example #A.1

adaptation, and learning. Fortunately, in these applications, we are rarely interested in evaluating the traditional complex derivative of $g(z)$. Instead, we are more interested in determining the location of the stationary points of $g(z)$. To do so, it is sufficient to rely on a different notion of differentiation based on what is sometimes known as the Wirtinger calculus [47, 252, 265], which we describe next.

∎

Scalar Arguments

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Scalar Arguments

We continue with the case in which $z \in \mathbb{C}$ is a scalar and allow $g(z)$ to be real or complex-valued so that $g(z) \in \mathbb{C}$. We again express z in terms of its real and imaginary parts as in (A.1), and similarly express $g(z)$ as a function of both x and y, i.e., as $g(x, y)$. The (Wirtinger) partial derivatives of $g(z)$ with respect to the complex arguments z and z^*, which we shall also refer to as the complex gradients of $g(z)$, are defined in terms of the partial derivatives of $g(x, y)$ with respect to the real arguments x and y as follows:

Scalar Arguments

$$\begin{cases} \dfrac{\partial g(z)}{\partial z} \triangleq \dfrac{1}{2}\left\{\dfrac{\partial g(x,y)}{\partial x} - j\,\dfrac{\partial g(x,y)}{\partial y}\right\} \\[2ex] \dfrac{\partial g(z)}{\partial z^*} \triangleq \dfrac{1}{2}\left\{\dfrac{\partial g(x,y)}{\partial x} + j\,\dfrac{\partial g(x,y)}{\partial y}\right\} \end{cases} \qquad \text{(A.9)}$$

Scalar Arguments

The above expressions can be grouped together in vector form as:

$$\begin{bmatrix} \partial g(z)/\partial z \\ \\ \partial g(z)/\partial z^* \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix}\begin{bmatrix} \partial g(x,y)/\partial x \\ \\ \partial g(x,y)/\partial y \end{bmatrix} \qquad \text{(A.10)}$$

so that, by inversion, it also holds that

$$\begin{bmatrix} \partial g(x,y)/\partial x \\ \\ \partial g(x,y)/\partial y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}\begin{bmatrix} \partial g(z)/\partial z \\ \\ \partial g(z)/\partial z^* \end{bmatrix} \qquad \text{(A.11)}$$

Scalar Arguments

The reason why the partial derivatives (A.9) are useful can be readily seen when $g(z)$ is real-valued, namely, $g(z) \in \mathbb{R}$. In that case, and by definition, a point $z^o = x^o + jy^o$ is said to be a stationary point of $g(z)$ if, and only if, (x^o, y^o) is a stationary point of $g(x, y)$. The latter condition is equivalent to requiring

$$\left.\frac{\partial g(x,y)}{\partial x}\right|_{x=x^o, y=y^o} = 0 \qquad \text{(A.12)}$$

and

$$\left.\frac{\partial g(x,y)}{\partial y}\right|_{x=x^o, y=y^o} = 0 \qquad \text{(A.13)}$$

Scalar Arguments

These two conditions combined are turn is equivalent to the following single condition in terms of the complex gradient vector:

$$\left.\frac{\partial g(z)}{\partial z}\right|_{z=z^o} = 0 \tag{A.14}$$

In this way, either of the partial derivatives defined by (A.9) enable us to locate stationary points of the real-valued function $g(z)$. Note that we are using the superscript notation "o", as in z^o, to refer to stationary points.

Example #A.2

Example A.2 (Wirtinger complex differentiation). We illustrate the definition of the partial derivatives (A.9) by considering a few examples. We will observe from the results in these examples that (Wirtinger) complex differentiation with respect to z treats z^* as a constant and, similarly, complex differentiation with respect to z^* treats z as a constant:

(1) Let $g(z) = z^2$. Then, $g(x,y) = (x^2 - y^2) + j2xy$ so that from (A.9):

$$\frac{\partial g(z)}{\partial z} = \frac{1}{2}(4x + j4y) = 2z, \qquad \frac{\partial g(z)}{\partial z^*} = 0 \tag{A.15}$$

Example #A.2

(2) Let $g(z) = |z|^2$. Then, $g(x, y) = x^2 + y^2$ and

$$\frac{\partial g(z)}{\partial z} = (x - jy) = z^*, \qquad \frac{\partial g(z)}{\partial z^*} = (x + jy) = z \tag{A.16}$$

(3) Let $g(z) = \kappa + \alpha z + \beta z^* + \gamma |z|^2$, where $(\kappa, \alpha, \beta, \gamma)$ are scalar constants. Then,

$$\frac{\partial g(z)}{\partial z} = \alpha + \gamma z^*, \qquad \frac{\partial g(z)}{\partial z^*} = \beta + \gamma z \tag{A.17}$$

$\blacksquare$

Vector Arguments

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Vector Arguments

We consider next the case in which z is a *column* vector argument, say, of size $M \times 1$, and whose individual entries are denoted by $\{z_m\}$, i.e.,

$$z = \text{col}\{z_1, z_2, \ldots, z_M\} \in \mathbb{C}^M \tag{A.18}$$

We continue to allow $g(z)$ to be real or complex-valued so that $g(z) \in \mathbb{C}$. The (Wirtinger) partial derivative of $g(z)$ with respect to z is again denoted by $\partial\, g(z)/\partial z$ and is defined as the *row* vector:

$$\frac{\partial g(z)}{\partial z} \triangleq \left[\begin{array}{cccc} \dfrac{\partial g}{\partial z_1} & \dfrac{\partial g}{\partial z_2} & \ldots & \dfrac{\partial g}{\partial z_M} \end{array}\right], \quad \left\{\begin{array}{l} z \text{ is a column} \\ \partial g/\partial z \text{ is a row} \end{array}\right. \tag{A.19}$$

Vector Arguments

in terms of the individual (Wirtinger) partial derivatives $\{\partial g/\partial z_m\}$. Expression (A.19) for $\partial\, g(z)/\partial z$ is also known as the *Jacobian* of $g(z)$. We shall refer to (A.19) as the complex gradient of $g(z)$ with respect to z and denote it more frequently by the alternative notation $\nabla_z\, g(z)$, i.e.,

$$\nabla_z\, g(z) \triangleq \left[\begin{array}{cccc} \dfrac{\partial g}{\partial z_1} & \dfrac{\partial g}{\partial z_2} & \ldots & \dfrac{\partial g}{\partial z_M} \end{array}\right], \quad \left\{\begin{array}{l} z \text{ is a column} \\ \nabla_z\, g(z) \text{ is a row} \end{array}\right. \tag{A.20}$$

Vector Arguments

There is not a clear convention in the literature on whether the gradient vector relative to z should be defined as a row vector (as in (A.20)) or as a column vector; both choices are common and both choices are useful. We prefer to use the *row* convention (A.20) because it leads to differentiation results that are consistent with what we are familiar with from the rules of traditional differentiation in the real domain — see Example A.3 below. This is largely a matter of convenience.

Vector Arguments

Likewise, along with (A.20), we define the complex gradient of $g(z)$ with respect to z^* to be the *column* vector:

$$\nabla_{z^*}\, g(z) \;\stackrel{\Delta}{=}\; \begin{bmatrix} \partial g/\partial z_1^* \\ \partial g/\partial z_2^* \\ \vdots \\ \partial g/\partial z_M^* \end{bmatrix} \equiv \frac{\partial g(z)}{\partial z^*}, \qquad \begin{cases} z^* \text{ is a row} \\ \nabla_{z^*}\, g(z) \text{ is a column} \end{cases} \tag{A.21}$$

Vector Arguments

Observe again the useful conclusion that when $g(z)$ is real-valued, then a vector $z^o = x^o + jy^o$ is a stationary point of $g(z)$ if, and only if,

$$\nabla_z \, g(z)|_{z=z^o} \;=\; 0 \tag{A.22}$$

Example #A.3

Example A.3 (Complex gradients). Let us again consider a few examples:

(1) Let $g(z) = a^* z$, where $\{a, z\}$ are column vectors. Then,

$$\nabla_z\, g(z) \;=\; a^*, \qquad \nabla_{z^*}\, g(z) \;=\; 0 \tag{A.23}$$

(2) Let $g(z) = \|z\|^2 = z^* z$, where z is a column vector. Then,

$$\nabla_z\, g(z) \;=\; z^*, \qquad \nabla_{z^*}\, g(z) \;=\; z \tag{A.24}$$

Example #A.3

(3) Let $g(z) = \kappa + a^*z + z^*b + z^*Cz$, where κ is a scalar, $\{a, b\}$ are column vectors, and C is a matrix. Then,

$$\nabla_z \, g(z) \;=\; a^* + z^*C, \qquad \nabla_{z^*} \, g(z) \;=\; b + Cz \tag{A.25}$$

$\blacksquare$

Real Arguments

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Real Arguments

When $z \in \mathbb{R}^M$ is real-valued and the function $g(z) \in \mathbb{R}$ is real-valued as well, the gradient vector is still defined as the row vector:

$$\nabla_z\, g(z) \triangleq \begin{bmatrix} \dfrac{\partial g}{\partial z_1} & \dfrac{\partial g}{\partial z_2} & \dots & \dfrac{\partial g}{\partial z_M} \end{bmatrix}, \quad \begin{cases} z \text{ is a column} \\ \nabla_z\, g(z) \text{ is a row} \end{cases} \tag{A.26}$$

in terms of the traditional partial derivatives of $g(z)$ with respect to the real scalar arguments $\{z_m\}$. Likewise, and in a manner that is consistent with (A.21), we define the gradient vector of $g(z)$ with respect to z^{T} to be the following *column* vector:

Real Arguments

$$\nabla_{z^{\mathsf{T}}}\, g(z) \;\triangleq\; \begin{bmatrix} \partial g/\partial z_1 \\ \partial g/\partial z_2 \\ \vdots \\ \partial g/\partial z_M \end{bmatrix}, \qquad \begin{cases} z^{\mathsf{T}} \text{ is a row} \\ \nabla_{z^{\mathsf{T}}}\, g(z) \text{ is a column} \end{cases} \tag{A.27}$$

In particular, note the useful relation

$$\nabla_{z^{\mathsf{T}}}\, g(z) \;=\; \left[\nabla_z\, g(z)\right]^{\mathsf{T}} \tag{A.28}$$

This relation holds for both cases when z itself is real-valued or complex-valued.

Example #A.4

Example A.4 (Quadratic cost functions I). Consider the quadratic function

$$g(z) = \kappa + a^{\mathsf{T}} z + z^{\mathsf{T}} b + z^{\mathsf{T}} C z \qquad \text{(A.29)}$$

where κ is a scalar, $\{a, b\}$ are column vectors of dimension $M \times 1$ each, and C is an $M \times M$ *symmetric* matrix (all of them are real-valued in this case). Then, it can be easily verified that

$$\nabla_z\, g(z) = a^{\mathsf{T}} + b^{\mathsf{T}} + 2z^{\mathsf{T}} C \qquad \text{(A.30)}$$

Example #A.4

The reason for the additional factor of two in the rightmost term can be justified by carrying out the calculation of the gradient vector explicitly. Indeed, if we denote the individual entries of $\{a, b, z, C\}$ by $\{a_m, b_m, z_m, C_{mn}\}$, then

$$g(z) \;=\; \kappa \;+\; \sum_{m=1}^{M}(a_m + b_m)z_m \;+\; \sum_{m=1}^{M}\sum_{n=1}^{M} z_m C_{mn} z_n \qquad \text{(A.31)}$$

Example #A.4

so that

$$\frac{\partial g(z)}{\partial z_m} = (a_m + b_m) + 2C_{mm}z_m + \sum_{n \neq m}^{M} (C_{mn} + C_{nm})z_n$$

$$= (a_m + b_m) + 2\sum_{n=1}^{M} C_{nm}z_n \qquad (A.32)$$

where we used the fact that C is symmetric and, hence, $C_{mn} = C_{nm}$. Collecting all the partial derivatives into the gradient vector defined by (A.26) we arrive at (A.30).

Example #A.4

Observe that while in the complex case, the arguments z and z^* are treated independently of each other during differentiation, this is not the case for the arguments z and z^T in the real case. In particular, since we can express the inner product $z^\mathsf{T}b$ as $b^\mathsf{T}z$, then the derivative of $z^\mathsf{T}b$ with respect to z is equal to the derivative of $b^\mathsf{T}z$ with respect to z (which explains the appearance of the term b^T in (A.30)).

■

Example #A.5

Example A.5 (Quadratic cost functions II). Consider the same quadratic function (A.29) with the only difference being that C is now arbitrary and *not* necessarily symmetric. Then, the same argument from Example A.4 will show that:

$$\nabla_z \, g(z) \;=\; a^{\mathsf{T}} + b^{\mathsf{T}} + z^{\mathsf{T}}(C + C^{\mathsf{T}}) \qquad \text{(A.33)}$$

where $2C$ in (A.30) is replaced by $C + C^{\mathsf{T}}$.

■

End of Lecture

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.