



# INFERENCE OVER NETWORKS

## LECTURE #6: Mean-Value Theorems

Professor Ali H. Sayed  
UCLA Electrical Engineering



# References



## Appendix D (Mean-Value Theorems, pp. 744-748):

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  $g(z) \in \mathbb{R}$  denote a *real-valued* function of a possibly vector argument  $z$ . We assume that  $g(z)$  is differentiable whenever necessary. In this appendix, we review useful integral equalities that involve increments of the function  $g(z)$  and increments of its gradient vector; the equalities correspond to extensions of the classical mean-value theorem from single-variable real calculus to the case of functions of several and possibly complex variables. We shall use the results of this appendix to establish useful bounds on the increments of strongly convex functions later in [Appendix E](#). We again treat both cases of real and complex arguments.

# Real 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.



# Real Arguments

Consider first the case in which the argument  $z \in \mathbb{R}^M$  is real-valued. We pick any  $M$ -dimensional vectors  $z_o$  and  $\Delta z$  and introduce the following real-valued and differentiable function of the scalar variable  $t \in [0, 1]$ :

$$f(t) \triangleq g(z_o + t \Delta z) \quad (\text{D.1})$$



# Real Arguments

Then, it holds that

$$f(0) = g(z_o), \quad f(1) = g(z_o + \Delta z) \quad (\text{D.2})$$

Using the fundamental theorem of calculus (e.g., [36, 151]) we have:

$$f(1) - f(0) = \int_0^1 \frac{df(t)}{dt} dt \quad (\text{D.3})$$

It further follows from definition (D.1) that

$$\frac{df(t)}{dt} = \frac{d}{dt} [g(z_o + t \Delta z)] = [\nabla_z g(z_o + t \Delta z)] \Delta z \quad (\text{D.4})$$



# Real Arguments

in terms of the inner product computation on the far right, where  $\nabla_z g(z)$  denotes the (row) gradient vector of  $g(z)$  with respect to  $z$ . Substituting (D.4) into (D.3) we arrive at the first desired mean-value theorem result (see, e.g., [191]):

$$g(z_o + \Delta z) - g(z_o) = \left( \int_0^1 \nabla_z g(z_o + t\Delta z) dt \right) \Delta z \quad (\text{D.5})$$



# Real Arguments

This result is a useful equality and it holds for any differentiable (*not necessarily convex*) real-valued function  $g(z)$ . The expression on the right-hand side is an inner product between the column vector  $\Delta z$  and the result of the integration, which is a row vector. Expression (D.5) tells us how the increment of the function  $g(z)$  in moving from  $z = z_o$  to  $z = z_o + \Delta z$  is related to the integral of the gradient vector of  $g(z)$  over the segment  $z_o + t \Delta z$  as  $t$  varies over the interval  $t \in [0, 1]$ .





# Real Arguments

We can derive a similar relation for increments of the gradient vector itself. To do so, we introduce the column vector function  $h(z) = \nabla_z^\top g(z)$  and apply (D.5) to its individual entries to conclude that

$$h(z_o + \Delta z) - h(z_o) = \left( \int_0^1 \nabla_z h(z_o + r \Delta z) dr \right) \Delta z \quad (\text{D.6})$$

Replacing  $h(z)$  by its definition, and transposing both sides of the above equality, we arrive at another useful mean-value theorem result:

$$\nabla_z g(z_o + \Delta z) - \nabla_z g(z_o) = \Delta z^\top \left( \int_0^1 \nabla_z^2 g(z_o + r \Delta z) dr \right) \quad (\text{D.7})$$



# Real Arguments

This expression tells us how increments in the gradient vector in moving from  $z = z_o$  to  $z = z_o + \Delta z$  are related to the integral of the Hessian matrix of  $g(z)$  over the segment  $z_o + r \Delta z$  and  $r$  varies over the interval  $r \in [0, 1]$ . In summary, we arrive at the following statement.



# Real Arguments

---

**Lemma D.1** (Mean-value theorem: Real arguments). Consider a real-valued and twice-differentiable function  $g(z) \in \mathbb{R}$ , where  $z \in \mathbb{R}^M$  is real-valued. Then, for any  $M$ -dimensional vectors  $z_o$  and  $\Delta z$ , the following increment equalities hold:

$$g(z_o + \Delta z) - g(z_o) = \left( \int_0^1 \nabla_z g(z_o + t \Delta z) dt \right) \Delta z \quad (\text{D.8})$$

$$\nabla_z g(z_o + \Delta z) - \nabla_z g(z_o) = (\Delta z)^\top \left( \int_0^1 \nabla_z^2 g(z_o + r \Delta z) dr \right) \quad (\text{D.9})$$

---

# Complex 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.



# Complex Arguments

13

Lecture #6: Mean-Value Theorems

EE210B: Inference over Networks (A. H. Sayed)

We now extend results (D.8) and (D.9) to the case when  $z \in \mathbb{C}^M$  is complex valued. The extension can be achieved by replacing  $z = x + jy$  by its real and imaginary parts  $\{x, y\}$ , applying results (D.8) and (D.9) to the resulting function  $g(v)$  of the  $2M \times 1$  extended real variable

$$v = \text{col}\{x, y\} \quad (\text{D.10})$$

and then transforming back to the complex domain. Indeed, as remarked earlier in (C.25), it is straightforward to verify that the vector  $v$  is related to the vector

$$u \triangleq \text{col}\{z, (z^*)^\top\} \quad (\text{D.11})$$



# Complex Arguments

14

Lecture #6: Mean-Value Theorems

EE210B: Inference over Networks (A. H. Sayed)

as follows:

$$\left\{ \begin{array}{l} \underbrace{\begin{bmatrix} z \\ (z^*)^\top \end{bmatrix}}_{\triangleq u} = \underbrace{\begin{bmatrix} I_M & jI_M \\ I_M & -jI_M \end{bmatrix}}_{\triangleq D} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\triangleq v} \\ \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_v = \underbrace{\frac{1}{2} \begin{bmatrix} I_M & I_M \\ -jI_M & jI_M \end{bmatrix}}_{=\frac{1}{2}D^*} \underbrace{\begin{bmatrix} z \\ (z^*)^\top \end{bmatrix}}_u \end{array} \right. \quad (\text{D.12})$$



# Complex Arguments

15

Lecture #6: Mean-Value Theorems

EE210B: Inference over Networks (A. H. Sayed)

or, more compactly,

$$u = Dv \quad \text{and} \quad v = \frac{1}{2}D^*u \quad (\text{D.13})$$

where we used the fact from (B.28) that  $DD^* = 2I_{2M}$ . We can now apply (D.8) to  $g(v)$  to get

$$g(v_o + \Delta v) - g(v_o) = \left( \int_0^1 \nabla_v g(v_o + t\Delta v) dt \right) \Delta v \quad (\text{D.14})$$



# Complex Arguments

where  $\nabla_v g(v)$  denotes the gradient vector of  $g(v)$ . We can rewrite (D.14) in terms of the original complex variables  $\{z_o, \Delta z\}$ . To do so, we call upon relation (C.32) and the equality  $g(z) = g(v)$  to rewrite (D.14) as

$$g(z_o + \Delta z) - g(z_o) = \tag{D.15}$$

$$\stackrel{\text{(D.13)}}{=} \frac{1}{2} \left( \int_0^1 \nabla_v g(v_o + t \Delta v) dt \right) \underbrace{D^* D \Delta v}_{\stackrel{\Delta}{=} \Delta u}$$

$$\stackrel{\text{(C.32)}}{=} \left( \int_0^1 \begin{bmatrix} \nabla_z g(z_o + t \Delta z) & (\nabla_{z^*} g(z_o + t \Delta z))^T \end{bmatrix} dt \right) \begin{bmatrix} \Delta z \\ (\Delta z^*)^T \end{bmatrix}$$





# Complex Arguments

17

Lecture #6: Mean-Value Theorems

EE210B: Inference over Networks (A. H. Sayed)

We then arrive at the desired mean-value theorem result in the complex case:

$$g(z_o + \Delta z) - g(z_o) = 2\text{Re} \left\{ \left( \int_0^1 \nabla_z g(z_o + t\Delta z) dt \right) \Delta z \right\} \quad (\text{D.16})$$

where we used the fact that for real-valued functions  $g(z)$  it holds that

$$\nabla_{z^*} g(z) = [\nabla_z g(z)]^* \quad (\text{D.17})$$

Expression (D.16) is the extension of (D.8) to the complex case.



# Complex Arguments

Similarly, applying (D.6) to  $h(v) = \nabla_{v^\top} g(v)$  we obtain that for any  $v_o$  and  $\Delta v$ :

$$\nabla_{v^\top} g(v_o + \Delta v) - \nabla_{v^\top} g(v_o) = \left( \int_0^1 \nabla_v^2 g(v_o + r \Delta v) dr \right) \Delta v \quad (\text{D.18})$$

Multiplying from the left by  $\frac{1}{2}D$  and using (C.30)–(C.31), as well as the fact that  $\frac{1}{4}DH_v(v)D^* = H_c(u)$  (recall (B.26)), we find that relation (D.18) defined in terms of  $\{v_o, \Delta v\}$  can be transformed into the mean-value theorem relation (D.20) in terms of the variables  $\{z_o, \Delta z\}$ .



# Complex Arguments

Expression (D.20) is the extension of (D.9) to the complex case. Observe how both gradient vectors relative to  $z^*$  and  $z^\top$  now appear in the relation. We show below in Example D.1 how the relation can be simplified in the special case when the Hessian matrix turns out to be block diagonal. In summary, we arrive at the following result.



# Complex Arguments

**Lemma D.2** (Mean-value theorem: Complex arguments). Consider a real-valued and twice-differentiable function  $g(z) \in \mathbb{R}$ , where  $z \in \mathbb{C}^M$  is complex-valued. Then, for any  $M$ -dimensional vectors  $z_o$  and  $\Delta z$ , the following increment equalities hold:

$$g(z_o + \Delta z) - g(z_o) = 2\text{Re} \left\{ \left( \int_0^1 \nabla_z g(z_o + t \Delta z) dt \right) \Delta z \right\} \quad (\text{D.19})$$

$$\begin{bmatrix} \nabla_{z^*} g(z_o + \Delta z) \\ \nabla_{z^\top} g(z_o + \Delta z) \end{bmatrix} - \begin{bmatrix} \nabla_{z^*} g(z_o) \\ \nabla_{z^\top} g(z_o) \end{bmatrix} = \left( \int_0^1 \nabla_z^2 g(z_o + r \Delta z) dr \right) \begin{bmatrix} \Delta z \\ (\Delta z^*)^\top \end{bmatrix} \quad (\text{D.20})$$



# Example #D.1

**Example D.1** (Block diagonal Hessian matrix). Consider the real-valued quadratic function

$$g(z) = \kappa + a^*z + z^*a + z^*Cz \quad (\text{D.21})$$

where  $\kappa$  is a real scalar,  $a$  is a column vector, and  $C$  is a Hermitian matrix. Then, the Hessian matrix of  $g(z)$  is block diagonal and given by

$$\nabla_z^2 g(z) \equiv H_c(u) = \begin{bmatrix} C & 0 \\ 0 & C^\top \end{bmatrix} \quad (\text{D.22})$$



# Example #D.1

In this case, expression (D.20) decouples into two separate and equivalent relations. Keeping one of the relations we get

$$\nabla_z g(z_o + \Delta z) = \nabla_z g(z_o) + (\Delta z)^* C \quad (\text{D.23})$$

Obviously, in this case, this relation could have been deduced directly from expression (D.21) by using the fact that

$$\nabla_z g(z) = a^* + z^* C \quad (\text{D.24})$$



# End of Lecture

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.