

# INFERENCE OVER NETWORKS

## LECTURE #7: Lipschitz Conditions

Professor Ali H. Sayed  
UCLA Electrical Engineering



# References



## Appendix E (Lipschitz Conditions, pp. 749-760):

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*  $\nu$ -strongly convex function of a possibly vector argument  $z$ . We assume that  $g(z)$  is differentiable whenever necessary. In this appendix, we use the mean-value theorems from [Appendix D](#) to derive some useful bounds on the increments of strongly convex functions. These bounds will assist in analyzing the mean-square-error stability and performance of distributed algorithms. We treat both cases of real and complex arguments.

# PERTURBATION BOUNDS

## 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.



# Last Time

---

**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:

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

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

---



# Perturbation Bounds

Consider first the case in which the argument  $z \in \mathbb{R}^M$  is real-valued. Let  $z^o$  denote the location of the unique global minimizer of  $g(z)$  so that  $\nabla_z g(z^o) = 0$ . Combining the mean-value theorem results (D.8) and (D.9) we get

$$g(z^o + \Delta z) - g(z^o) = (\Delta z)^\top \left[ \int_0^1 \int_0^1 t \nabla_z^2 g(z^o + tr \Delta z) dr dt \right] \Delta z \quad (\text{E.1})$$

# Real Domain



Now assume the Hessian matrix of  $g(z)$  is uniformly bounded from above, i.e.,

$$\nabla_z^2 g(z) \leq \delta I_M, \quad \text{for all } z \quad (\text{E.2})$$

and for some  $\delta > 0$ . It follows from (E.1) that

$$g(z^o + \Delta z) - g(z^o) \leq \frac{\delta}{2} \|\Delta z\|^2 \quad (\text{E.3})$$

which leads to the following useful statement for strongly-convex functions.



# Statement

**Lemma E.1** (Perturbation bound: Real arguments). Consider a  $\nu$ -strongly convex and twice-differentiable function  $g(z) \in \mathbb{R}$  and let  $z^o \in \mathbb{R}^M$  denote its global minimizer. Assume that its  $M \times M$  Hessian matrix (defined according to the first row in Table B.1 or equation (B.29)) is uniformly bounded from above by  $\nabla_z^2 g(z) \leq \delta I_M$ , for all  $z$  and for some  $\delta > 0$ . We already know from item (c) in (C.18) that the same Hessian matrix is uniformly bounded from below by  $\nu I_M$ , i.e.,

$$\nu I_M \leq \nabla_z^2 g(z) \leq \delta I_M, \quad \text{for all } z \quad (\text{E.4})$$

Under condition (E.4), it follows from (C.16) and (E.3) that, for any  $\Delta z$ , the function increments are bounded by the squared Euclidean norm of  $\Delta z$  as follows:

$$\frac{\nu}{2} \|\Delta z\|^2 \leq g(z^o + \Delta z) - g(z^o) \leq \frac{\delta}{2} \|\Delta z\|^2 \quad (\text{E.5})$$



# Real Domain



One useful conclusion that follows from (E.5) is that under condition (E.4), every strongly convex function  $g(z)$  can be sandwiched between two quadratic functions, namely,

$$g(z^o) + \frac{\nu}{2} \|z - z^o\|^2 \leq g(z) \leq g(z^o) + \frac{\delta}{2} \|z - z^o\|^2 \quad (\text{E.6})$$

# Real Domain



A second useful conclusion can be deduced from (E.1) when the size of  $\Delta z$  is small and when the Hessian matrix of  $g(z)$  is smooth enough in a small neighborhood around  $z = z^o$ . Specifically, assume the Hessian matrix function is locally Lipschitz continuous in a small neighborhood around  $z = z^o$ , namely,

$$\left\| \nabla_z^2 g(z^o + \Delta z) - \nabla_z^2 g(z^o) \right\| \leq \kappa \|\Delta z\| \quad (\text{E.7})$$

for sufficiently small values  $\|\Delta z\| \leq \epsilon$  and for some  $\kappa > 0$ . This condition implies that we can write

$$\nabla_z^2 g(z^o + \Delta z) = \nabla_z^2 g(z^o) + O(\|\Delta z\|) \quad (\text{E.8})$$

# Real Domain



It then follows from equality (E.1) that, for sufficiently small  $\Delta z$ :

$$\begin{aligned} g(z^o + \Delta z) - g(z^o) &= (\Delta z)^\top \left[ \frac{1}{2} \nabla_z^2 g(z^o) \right] \Delta z + O(\|\Delta z\|^3) \\ &\approx (\Delta z)^\top \left[ \frac{1}{2} \nabla_z^2 g(z^o) \right] \Delta z \\ &= \|\Delta z\|_{\frac{1}{2} \nabla_z^2 g(z^o)}^2 \end{aligned} \tag{E.9}$$

where the symbol  $\approx$  in the second line is used to indicate that higher-order powers in  $\|\Delta z\|$  are being ignored. Moreover, for any Hermitian positive-definite weighting matrix  $W > 0$ , the notation  $\|x\|_W^2$  refers to the weighted square Euclidean norm  $x^* W x$ .

# Real Domain



We conclude from (E.9) that the increment in the value of the function in a small neighborhood around  $z = z^o$  can be well approximated by means of a weighted square Euclidean norm; the weighting matrix in this case is equal to the Hessian matrix of  $g(z)$  evaluated at  $z = z^o$  and scaled by  $1/2$ . The error in this approximate evaluation is in the order of  $\|\Delta z\|^3$ .



# Statement

---

**Lemma E.2** (Perturbation approximation: Real arguments). Consider the same setting of [Lemma E.1](#) and assume additionally that the Hessian matrix function is locally Lipschitz continuous in a small neighborhood around  $z = z^o$  as defined by [\(E.7\)](#). It then follows that the increment in the value of the function  $g(z)$  for sufficiently small variations around  $z = z^o$  can be well approximated by

$$g(z^o + \Delta z) - g(z^o) \approx \Delta z^\top \left[ \frac{1}{2} \nabla_z^2 g(z^o) \right] \Delta z \quad (\text{E.10})$$

where the approximation error is in the order of  $O(\|\Delta z\|^3)$ .

---



# Example #E.1

**Example E.1** (Quadratic cost functions with real arguments). Consider a quadratic function of the form

$$g(z) = \kappa - a^\top z - z^\top a + z^\top C z \quad (\text{E.11})$$

where  $\kappa$  is a scalar,  $a$  is a column vector, and  $C$  is a symmetric positive-definite matrix. It is straightforward to verify, by expanding the right-hand side in the expression below, that  $g(z)$  can also be written as

$$g(z) = \kappa - a^\top C^{-1} a + (z - C^{-1} a)^\top C (z - C^{-1} a) \quad (\text{E.12})$$



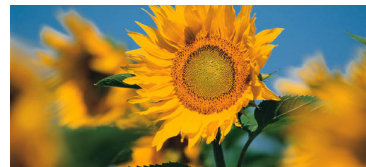
# Example #E.1

The Hessian matrix is  $\nabla_z^2 g(z) = 2C$  and it is clear that

$$2\lambda_{\min}(C) I_M \leq \nabla_z^2 g(z) \leq 2\lambda_{\max}(C) I_M \quad (\text{E.13})$$

in terms of the smallest and largest eigenvalues of  $C$ , which are both positive. Therefore, condition (E.4) is automatically satisfied with

$$\nu = 2\lambda_{\min}(C), \quad \delta = 2\lambda_{\max}(C) \quad (\text{E.14})$$



# Example #E.1

Likewise, condition (E.7) is obviously satisfied since the Hessian matrix in this case is constant and independent of  $z$ . The function  $g(z)$  has a unique global minimizer and it occurs at the point  $z = z^o$  where  $\nabla_z g(z^o) = 0$ . We know from the expression for  $g(z)$  that

$$\nabla_z g(z) = -2a^\top + 2z^\top C \quad (\text{E.15})$$

so that  $z^o = C^{-1}a$  and

$$g(z^o) = \kappa - a^\top C^{-1}a \quad (\text{E.16})$$





# Example #E.1

Therefore, applying (E.6) we conclude that

$$g(z^o) + \lambda_{\min}(C) \|z - C^{-1}a\|^2 \leq g(z) \leq g(z^o) + \lambda_{\max}(C) \|z - C^{-1}a\|^2 \quad (\text{E.17})$$

Note that we could have arrived at this result directly from (E.12) as well.

Moreover, from result (E.10) we would estimate that, for sufficiently small  $\|\Delta z\|$ ,

$$g(z^o + \Delta z) - g(z^o) \approx \|\Delta z\|_C^2 \quad (\text{E.18})$$

Actually, in this case, exact equality holds in (E.18) for *any*  $\Delta z$  due to the quadratic nature of the function  $g(z)$ . Indeed, note from (E.12) that

$$g(z) = g(z^o) + \|z - z^o\|_C^2 \quad (\text{E.19})$$



# Example #E.1

so that if we set  $z = z^o + \Delta z$ , for *any*  $\Delta z$ , the above relation gives

$$g(z^o + \Delta z) - g(z^o) = \|\Delta z\|_C^2, \quad \text{for any } \Delta z \quad (\text{E.20})$$

which is a stronger result than (E.18); note in particular that  $\Delta z$  does not need to be infinitesimally small any more, as was the case with (E.10); this latter relation is useful for more general choices of  $g(z)$  that are not necessarily quadratic in  $z$ .



# LIPSCHITZ CONDITIONS

## 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.



# Lipschitz Conditions

20

Lecture #7: Lipschitz Conditions

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

The statement of [Lemma E.1](#) requires the Hessian matrix to be upper bounded as in [\(E.2\)](#), i.e.,  $\nabla_z^2 g(z) \leq \delta I_M$  for all  $z$ . For differentiable convex functions, this condition is equivalent to requiring the gradient vector to be Lipschitz continuous, i.e., to satisfy

$$\|\nabla_z g(z_2) - \nabla_z g(z_1)\| \leq \delta \|z_2 - z_1\| \quad (\text{E.21})$$

for all  $z_1$  and  $z_2$ .



# Lipschitz Conditions

Since it is customary in the literature to rely more frequently on Lipschitz conditions, the following statement establishes the equivalence of conditions (E.2) and (E.21) for general convex functions (that are not necessarily strongly-convex). One advantage of using condition (E.21) instead of (E.2) is that the function  $g(z)$  would not need to be twice-differentiable since condition (E.21) only involves the gradient vector of the function.



# Statement

---

**Lemma E.3** (Lipschitz and bounded Hessian matrix). Consider a real-valued and twice-differentiable convex function  $g(z) \in \mathbb{R}$ . Then, the following two conditions are equivalent:

$$\nabla_z^2 g(z) \leq \delta I_M, \text{ for all } z \iff \|\nabla_z g(z_2) - \nabla_z g(z_1)\| \leq \delta \|z_2 - z_1\|, \text{ for all } z_1, z_2$$

(E.22)

---

# Proof



*Proof.* Assume first that the Hessian matrix,  $\nabla_z^2 g(z)$ , is uniformly upper bounded by  $\delta I_M$  for all  $z$ ; we know that it is nonnegative definite since  $g(z)$  is convex and, therefore,  $\nabla_z^2 g(z)$  is lower bounded by zero. We pick any  $z_1$  and  $z_2$  and introduce the column vector function  $h(z) = \nabla_{z^\top} g(z)$ . Applying (D.8) to  $h(z)$  gives

$$h(z_2) - h(z_1) = \left( \int_0^1 \nabla_z h(z_1 + t(z_2 - z_1)) dt \right) (z_2 - z_1) \quad (\text{E.23})$$

# Proof



so that using  $0 \leq \nabla_z^2 g(z) \leq \delta I_M$ , we get

$$\|\nabla_{z^\top} g(z_2) - \nabla_{z^\top} g(z_1)\| \leq \left( \int_0^1 \delta dt \right) \|z_2 - z_1\| \quad (\text{E.24})$$

and we arrive at the Lipschitz condition on the right-hand side of (E.22) since  $\nabla_{z^\top} g(z) = [\nabla_z g(z)]^\top$ .



# Proof



Conversely, assume the Lipschitz condition on the right-hand side of (E.22) holds, and introduce the column vector function  $f(t) = \nabla_{z^\top} g(z + t\Delta z)$  defined in terms of a scalar real parameter  $t$ . Then,

$$\frac{df(t)}{dt} = [\nabla_z^2 g(z + t\Delta z)] \Delta z \quad (\text{E.25})$$

Now, for any  $\Delta t$  and in view of the Lipschitz condition, it holds that

$$\begin{aligned} \|f(t + \Delta t) - f(t)\| &= \|\nabla_{z^\top} g(z + (t + \Delta t)\Delta z) - \nabla_{z^\top} g(z + t\Delta z)\| \\ &\leq \delta |\Delta t| \|\Delta z\| \end{aligned} \quad (\text{E.26})$$

# Proof



so that

$$\underbrace{\lim_{\Delta t \rightarrow 0} \frac{\|f(t + \Delta t) - f(t)\|}{|\Delta t|}}_{= \|df(t)/dt\|} \leq \delta \|\Delta z\| \quad (\text{E.27})$$

Using (E.25) we conclude that

$$\| [\nabla_z^2 g(z + t\Delta z)] \Delta z \| \leq \delta \|\Delta z\|, \quad \text{for any } t, z \text{ and } \Delta z \quad (\text{E.28})$$

# Proof



Setting  $t = 0$ , squaring both sides, and recalling that the Hessian matrix is symmetric, we obtain

$$(\Delta z)^\top [\nabla_z^2 g(z)]^2 \Delta z \leq \delta^2 \|\Delta z\|^2, \quad \text{for any } z, \Delta z \quad (\text{E.29})$$

from which we conclude that  $\nabla_z^2 g(z) \leq \delta I_M$  for all  $z$ , as desired.



# Local and Global Conditions



We can additionally verify that the local Lipschitz condition (E.7) used in Lemma E.2 is actually equivalent to a global Lipschitz property on the Hessian matrix under condition (E.4).



# Statement

---

**Lemma E.4** (Global Lipschitz condition). Consider a real-valued and twice-differentiable  $\nu$ -strongly convex function  $g(z) \in \mathbb{R}$  and assume it satisfies conditions (E.4) and (E.7). It then follows that the Hessian matrix of  $g(z)$  is globally Lipschitz relative to  $z^o$ , namely, it satisfies

$$\|\nabla_z^2 g(z) - \nabla_z^2 g(z^o)\| \leq \kappa' \|z - z^o\|, \quad \text{for all } z \quad (\text{E.30})$$

where the positive scalar  $\kappa'$  is defined in terms of the parameters  $\{\kappa, \delta, \nu, \epsilon\}$  as

$$\kappa' = \max \left\{ \kappa, \frac{\delta - \nu}{\epsilon} \right\} \quad (\text{E.31})$$

---

# Proof



*Proof.* Following [278], for any vector  $x$ , it holds that

$$\begin{aligned} x^\top [\nabla_z^2 g(z) - \nabla_z^2 g(z^o)] x &= x^\top \nabla_z^2 g(z) x - x^\top \nabla_z^2 g(z^o) x \\ &\stackrel{(\text{E.4})}{\leq} \delta \|x\|^2 - \nu \|x\|^2 \\ &= (\delta - \nu) \|x\|^2 \end{aligned} \tag{E.32}$$

And since the Hessian matrix difference is **symmetric**, we conclude that  $\nabla_z^2 g(z) - \nabla_z^2 g(z^o) \leq (\delta - \nu) I_M$  so that, in terms of the 2-induced norm:

$$\|\nabla_z^2 g(z) - \nabla_z^2 g(z^o)\| \leq \delta - \nu \tag{E.33}$$

# Proof



Now, consider any vector  $z$  such that  $\|z - z^o\| \leq \epsilon$ . Then,

$$\|\nabla_z^2 g(z) - \nabla_z^2 g(z^o)\| \stackrel{\text{(E.7)}}{\leq} \kappa \|z - z^o\| \stackrel{\text{(E.31)}}{\leq} \kappa' \|z - z^o\| \quad (\text{E.34})$$

On the other hand, for any vector  $z$  such that  $\|z - z^o\| > \epsilon$ , we have

$$\|\nabla_z^2 g(z) - \nabla_z^2 g(z^o)\| \stackrel{\text{(E.33)}}{\leq} \left( \frac{\delta - \nu}{\epsilon} \right) \epsilon \stackrel{\text{(E.31)}}{\leq} \kappa' \|z - z^o\| \quad (\text{E.35})$$

□

# PERTURBATION BOUNDS

## 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.





# Perturbation Bounds

The statement below extends the result of [Lemma E.1](#) to the case of complex arguments,  $z \in \mathbb{C}^M$ . Comparing the bounds in [\(E.37\)](#) with the earlier result [\(E.5\)](#), we observe that the relations are identical. The only difference in the complex case relative to the real case is that the upper and lower bounds on the complex Hessian matrix in [\(E.36\)](#) are scaled by  $1/2$  relative to the bounds in [\(E.4\)](#).



# Statement

**Lemma E.5** (Perturbation bound: Complex arguments). Consider a  $\nu$ -strongly convex and twice-differentiable function  $g(z) \in \mathbb{R}$  and let  $z^o \in \mathbb{C}^M$  denote its global minimizer. The function  $g(z)$  is real-valued but  $z$  is now complex-valued. Assume that the  $2M \times 2M$  complex Hessian matrix of  $g(z)$  (defined according to the last row of [Table B.1](#) and [\(B.29\)](#)) is uniformly bounded from above by  $\nabla_z^2 g(z) \leq \frac{\delta}{2} I_{2M}$ , for all  $z$  and for some  $\delta > 0$ . We already know from item (c) in [\(C.44\)](#) that the same Hessian matrix is uniformly bounded from below by  $\frac{\nu}{2} I_{2M}$ , i.e.,

$$\frac{\nu}{2} I_{2M} \leq \nabla_z^2 g(z) \leq \frac{\delta}{2} I_{2M}, \quad \text{for all } z \quad (\text{E.36})$$



# Statement

Under condition (E.36) it holds that, for any  $\Delta z$ , the function increments are bounded by the squared Euclidean norm of  $\Delta z$  as follows:

$$\frac{\nu}{2} \|\Delta z\|^2 \leq g(z^o + \Delta z) - g(z^o) \leq \frac{\delta}{2} \|\Delta z\|^2 \quad (\text{E.37})$$

---

# Proof



*Proof.* The argument is based on expressing  $z$  in terms of its real and imaginary parts,  $z = x + jy$ , transforming  $g(z)$  into a function of the  $2M \times 1$  extended real variable  $v = \text{col}\{x, y\}$ , and then applying the result of [Lemma E.1](#) to  $g(v)$ .

To begin with, recall that the  $2M \times 2M$  Hessian matrix of  $g(v)$  is denoted by  $H(v)$  and is constructed according to the second row of [Table B.1](#). This real Hessian matrix is related by [\(B.26\)](#) to the complex Hessian matrix,  $H_c(u)$ , of  $g(z)$  and which we are denoting by  $\nabla_z^2 g(z)$  in the statement of the lemma. Therefore, the upper bound on  $\nabla_z^2 g(z)$  in [\(E.36\)](#) can be transformed into an upper bound on  $H(v)$  by noting that

# Proof



$$H(v) \stackrel{\text{(B.26)}}{=} D^* [\nabla_z^2 g(z)] D \leq \frac{\delta}{2} D^* D = \delta I_{2M} \quad (\text{E.38})$$

since  $D^* D = 2I_{2M}$  and, hence,  $H(v) \leq \delta I_{2M}$ . Combining this result with (C.45) we conclude that the Hessian matrix  $H(v)$  is bounded as follows:

$$\nu I_{2M} \leq H(v) \leq \delta I_{2M} \quad (\text{E.39})$$

# Proof



Consequently, if we apply the result of [Lemma E.1](#) to the function  $g(v)$ , whose argument  $v$  is real, we find that

$$\frac{\nu}{2} \|\Delta v\|^2 \leq g(v^o + \Delta v) - g(v^o) \leq \frac{\delta}{2} \|\Delta v\|^2 \quad (\text{E.40})$$

which is equivalent to the desired relation ([E.37](#)) in terms of the original variables  $\{z^o, \Delta z\}$  since, for any  $z$ ,  $g(z) = g(v)$  and  $\|z\| = \|v\|$ .





# Consequence

One useful conclusion that follows from (E.37) is that under condition (E.36), the strongly convex function  $g(z)$  can be sandwiched between two quadratic functions, namely,

$$g(z^o) + \frac{\nu}{2} \|z - z^o\|^2 \leq g(z) \leq g(z^o) + \frac{\delta}{2} \|z - z^o\|^2 \quad (\text{E.41})$$

A second useful conclusion is an extension of (E.10) to the case of complex arguments  $z$ . Introduce the extended vector:

$$\Delta z^e \triangleq \begin{bmatrix} \Delta z \\ (\Delta z^*)^\top \end{bmatrix} \quad (\text{E.42})$$

# Statement



**Lemma E.6** (Perturbation approximation: Complex arguments). Consider the same setting of [Lemma E.5](#) and assume additionally that the Hessian matrix function is locally Lipschitz continuous in a small neighborhood around  $z = z^o$ , namely,

$$\|\nabla_z^2 g(z^o + \Delta z) - \nabla_z^2 g(z^o)\| \leq \kappa \|\Delta z\| \quad (\text{E.43})$$

for sufficiently small values  $\|\Delta z\| \leq \epsilon$  and for some  $\kappa > 0$ . It then follows that the increment in the value of the function  $g(z)$  for small variations around  $z = z^o$  can be well approximated by:

$$g(z^o + \Delta z) - g(z^o) \approx (\Delta z^e)^* \left[ \frac{1}{2} \nabla_z^2 g(z^o) \right] \Delta z^e \quad (\text{E.44})$$

where the approximation error is in the order of  $O(\|\Delta z\|^3)$ .



# Proof



*Proof.* Result (E.44) can be derived from (E.10) as follows. We again transform  $g(z)$  into the function  $g(v)$  of the real variable  $v = \text{col}\{x, y\}$  and then apply (E.10) to  $g(v)$  for sufficiently small  $\Delta v$ , which gives

$$g(v^o + \Delta v) - g(v^o) \approx (\Delta v)^\top \left[ \frac{1}{2} H(v^o) \right] \Delta v, \quad \text{as } \Delta v \rightarrow 0 \quad (\text{E.45})$$

in terms of the  $2M \times 2M$  Hessian matrix of  $g(v)$  evaluated at  $v = v^o$ . This Hessian matrix is related to the complex Hessian matrix  $H_c(u^o)$  according to (B.26). Thus, observe that

# Proof



$$\begin{aligned}
 (\Delta v)^\top \left[ \frac{1}{2} H(v^o) \right] \Delta v &= \frac{1}{4} (\Delta v)^\top D^* D \left[ \frac{1}{2} H(v^o) \right] D^* D \Delta v \\
 &\stackrel{\text{(D.13)}}{=} \frac{1}{2} \underbrace{(\Delta v)^\top D^*}_{(\Delta u)^*} \underbrace{\frac{1}{4} D H(v^o) D^*}_{H_c(u^o)} \underbrace{D \Delta v}_{\Delta u} \\
 &\stackrel{\text{(B.26)}}{=} \frac{1}{2} (\Delta u)^* H_c(u^o) \Delta u \\
 &\stackrel{\text{(B.24)}}{=} \frac{1}{2} \begin{bmatrix} (\Delta z)^* & \Delta z^\top \end{bmatrix} \nabla_z^2 g(z^o) \begin{bmatrix} \Delta z \\ (\Delta z^*)^\top \end{bmatrix} \\
 &= (\Delta z^e)^* \left[ \frac{1}{2} \nabla_z^2 g(z^o) \right] \Delta z^e \tag{E.46}
 \end{aligned}$$

as claimed.





# Example #E.2

**Example E.2** (Quadratic cost functions with complex arguments). Let us illustrate the above result by considering a quadratic function of the form

$$g(z) = \kappa - a^*z - z^*a + z^*Cz \quad (\text{E.47})$$

where  $\kappa$  is a scalar,  $a$  is a column vector, and  $C$  is a Hermitian positive-definite matrix. It is straightforward to verify, by expanding the right-hand side in the expression below, that  $g(z)$  can be also written as

$$g(z) = \kappa - a^*C^{-1}a + (z - C^{-1}a)^*C(z - C^{-1}a) \quad (\text{E.48})$$



# Example #E.2

The Hessian matrix in this case is  $2M \times 2M$  and given by

$$\nabla_z^2 g(z) = \begin{bmatrix} C & 0 \\ 0 & C^\top \end{bmatrix} \quad (\text{E.49})$$

It is clear that

$$\lambda_{\min}(C) I_{2M} \leq \nabla_z^2 g(z) \leq \lambda_{\max}(C) I_{2M} \quad (\text{E.50})$$

in terms of the smallest and largest eigenvalues of  $C$ , which are both positive. Therefore, condition (E.36) is automatically satisfied with

$$\nu = 2\lambda_{\min}(C), \quad \delta = 2\lambda_{\max}(C) \quad (\text{E.51})$$



# Example #E.2

Likewise, condition (E.43) is satisfied since the Hessian matrix is constant and independent of  $z$ . The function  $g(z)$  has a unique global minimizer and it occurs at the point  $z = z^o$  where  $\nabla_z g(z^o) = 0$ . We know from expression (E.48) for  $g(z)$  that  $z^o = C^{-1}a$  and  $g(z^o) = \kappa - a^*C^{-1}a$ . Therefore, applying (E.41) we conclude that

$$g(z^o) + \lambda_{\min}(C) \|z - C^{-1}a\|^2 \leq g(z) \leq g(z^o) + \lambda_{\max}(C) \|z - C^{-1}a\|^2 \quad (\text{E.52})$$

Note that we could have arrived at this result directly from (E.48) as well.



# Example #E.2

Moreover, we would estimate from (E.44) that

$$\begin{aligned} g(z^o + \Delta z) - g(z^o) &\approx \frac{1}{2} \begin{bmatrix} (\Delta z)^* & (\Delta z)^\top \end{bmatrix} \begin{bmatrix} C & 0 \\ 0 & C^\top \end{bmatrix} \begin{bmatrix} \Delta z \\ (\Delta z^*)^\top \end{bmatrix} \\ &= \|\Delta z\|_C^2 \end{aligned} \tag{E.53}$$

where the notation  $\|x\|_C^2$  now denotes the squared Euclidean quantity  $x^*Cx$ . Actually, in this case, exact equality holds in (E.53) for *any*  $\Delta z$  due to the quadratic nature of the function  $g(z)$ . Indeed, note that (E.48) can be rewritten as

$$g(z) = g(z^o) + \|z - z^o\|_C^2 \tag{E.54}$$



# Example #E.2

so that if we set  $z = z^o + \Delta z$ , for *any*  $\Delta z$ , the above relation gives

$$g(z^o + \Delta z) - g(z^o) = \|\Delta z\|_C^2, \quad \text{for any } \Delta z \quad (\text{E.55})$$

which is a stronger result than the approximation in (E.53); note in particular that  $\Delta z$  does not need to be infinitesimally small any more, as was the case with (E.44); this latter result is applicable to more general choices of  $g(z)$  that are not necessarily quadratic in  $z$ .



# LIPSCHITZ CONDITIONS

## 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.





# Lipschitz Conditions

49

Lecture #7: Lipschitz Conditions

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

The statement of [Lemma E.5](#) requires the Hessian matrix to be upper bounded as in [\(E.36\)](#), i.e.,  $\nabla_z^2 g(z) \leq \frac{\delta}{2} I_{2M}$  for all  $z$ . As was the case with real arguments in [Lemma E.3](#), we can argue that for general convex functions (that are not necessarily strongly convex), this condition is equivalent to requiring the gradient vector to be Lipschitz continuous.

# Statement



---

**Lemma E.7** (Lipschitz and bounded Hessian matrix). Consider a real-valued and twice-differentiable convex function  $g(z) \in \mathbb{R}$ , where  $z \in \mathbb{C}^M$  is now complex valued. Then, the following two conditions are equivalent:

$$\nabla_z^2 g(z) \leq \frac{\delta}{2} I_{2M}, \text{ for all } z \iff \|\nabla_z g(z_2) - \nabla_z g(z_1)\| \leq \frac{\delta}{2} \|z_2 - z_1\|, \text{ for all } z_1, z_2$$

---

(E.56)

# Proof



*Proof.* The above result can be derived from (E.22) as follows. We transform  $g(z)$  into the function  $g(v)$  of the real variable  $v = \text{col}\{x, y\}$ , where  $z = x + jy$ , and then apply (E.22) to  $g(v)$ .

First, recall from the argument that led to (E.39) that the complex Hessian matrix of  $g(z)$  is bounded by  $\frac{\delta}{2}I_{2M}$  if, and only if, the real Hessian matrix of  $g(v)$  is bounded by  $\delta I_{2M}$ . Using this observation and applying (E.22) to  $g(v)$  we get

# Proof



$$\begin{aligned}\nabla_z^2 g(z) &\leq \frac{\delta}{2} I_{2M} \stackrel{\text{(E.39)}}{\iff} \nabla_v^2 g(v) \leq \delta I_{2M}, \quad \text{for all } v \\ &\stackrel{\text{(E.22)}}{\iff} \|\nabla_v g(v_2) - \nabla_v g(v_1)\| \leq \delta \|v_2 - v_1\| \tag{E.57}\end{aligned}$$

for any  $v_1, v_2$ . Now we know from (C.32) that

$$\frac{1}{2} [\nabla_v g(v)] D^* = \begin{bmatrix} \nabla_z g(z) & (\nabla_{z^*} g(z))^T \end{bmatrix} \tag{E.58}$$

# Proof



Recalling from (B.28) that the matrix  $D^*/\sqrt{2}$  is unitary, we get

$$\|\nabla_v g(v_2) - \nabla_v g(v_1)\| = \quad (\text{E.59})$$

$$\begin{aligned} &= \left\| [\nabla_v g(v_2) - \nabla_v g(v_1)] \cdot \frac{D^*}{\sqrt{2}} \right\| \\ &\stackrel{(\text{E.58})}{=} \sqrt{2} \cdot \left\| \begin{bmatrix} \nabla_z g(z_2) - \nabla_z g(z_1) & (\nabla_{z^*} g(z_2) - \nabla_{z^*} g(z_1))^T \end{bmatrix} \right\| \\ &= 2 \cdot \|\nabla_z g(z_2) - \nabla_z g(z_1)\| \end{aligned}$$

# Proof



where we used (D.17). Noting that  $\|v_2 - v_1\| = \|z_2 - z_1\|$  and substituting into (E.57) we conclude that

$$\nabla_z^2 g(z) \leq \frac{\delta}{2} I_{2M} \iff \|\nabla_z g(z_2) - \nabla_z g(z_1)\| \leq \frac{\delta}{2} \|z_2 - z_1\|, \quad \text{for all } z_1, z_2 \quad (\text{E.60})$$

as claimed.



# Local and Global Conditions



We can again verify that the local Lipschitz condition (E.43) used in Lemma E.6 is equivalent to a global Lipschitz property on the Hessian matrix under the bounds (E.36). The proof of the following result is similar to that of Lemma E.4.



# Statement

---

**Lemma E.8** (Global Lipschitz condition). Consider a real-valued and twice-differentiable  $\nu$ -strongly convex function  $g(z) \in \mathbb{R}$  and assume it satisfies conditions (E.36) and (E.43). It then follows that the  $2M \times 2M$  Hessian matrix of  $g(z)$  is globally Lipschitz relative to  $z^o \in \mathbb{C}^M$ , namely,

$$\|\nabla_z^2 g(z) - \nabla_z^2 g(z^o)\| \leq \kappa' \|z - z^o\|, \quad \text{for all } z \quad (\text{E.61})$$

where the positive scalar  $\kappa'$  is defined in terms of the parameters  $\{\kappa, \delta, \nu, \epsilon\}$  as

$$\kappa' = \max \left\{ \kappa, \frac{\delta - \nu}{2\epsilon} \right\} \quad (\text{E.62})$$

---



# 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.