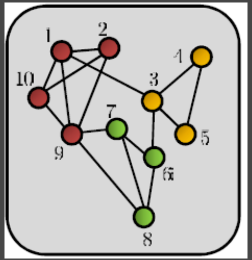

INFERENCE OVER NETWORKS

LECTURE #4: Convex Functions

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.

Reference

Appendix C (Convex Functions, pp. 730-743):

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, $z \in \mathbb{C}^M$. It is sufficient for our purposes to assume that $g(z)$ is differentiable whenever necessary (although we shall also comment on the situation in which $g(z)$ may not be differentiable at some points). By differentiability here we mean that the (Wirtinger) complex gradient vector, $\nabla_z\, g(z)$, and the Hessian matrix, $\nabla_z^2\, g(z)$, both exist in the manner defined in Appendices A and B.

Setting

In particular, if we express z in terms of its real and imaginary arguments, $z = x + jy$, then we are assuming that the following partial derivatives exist whenever necessary:

$$\frac{\partial g(x,y)}{\partial x_m}, \quad \frac{\partial g(x,y)}{\partial y_n}, \quad \frac{\partial^2 g(x,y)}{\partial x_m^2}, \quad \frac{\partial^2 g(x,y)}{\partial y_n}, \quad \frac{\partial^2 g(x,y)}{\partial x_m \partial y_n} \qquad \text{(C.1)}$$

for $n, m = 1, 2, \ldots, M$, and where $\{x_m, y_n\}$ denote the individual entries of the vectors $x, y \in \mathbb{R}^M$.

Setting

In the sequel, we define convexity for both cases when $z \in \mathbb{R}^M$ is real-valued and when $z \in \mathbb{C}^M$ is complex-valued. We start with the former case of real z, which is the situation most commonly studied in the literature [29, 45, 178, 191]. Subsequently, we explain how the definitions and results extend to functions of complex arguments, z; these extensions are necessary to deal with situations that arise in the context of adaptation and learning in signal processing and communications problems.

Real Domain

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.

Convexity in Real Domain

We assume initially that the argument $z \in \mathbb{R}^M$ is real-valued where, as already stated earlier, the function $g(z) \in \mathbb{R}$ is also real-valued. We discuss three forms of convexity: the standard definition of convexity followed by strict convexity and then strong convexity.

Convex Sets

We first introduce the notion of convex sets. A set $\mathcal{S} \subset \mathbb{R}^M$ is said to be convex if for any pair of points $z_1, z_2 \in \mathcal{S}$, all points that lie on the line segment connecting z_1 and z_2 also belong to $\mathcal{S}$. Specifically,

$$\forall z_1, z_2 \in \mathcal{S} \;\text{ and }\; 0 \leq \alpha \leq 1 \;\Longrightarrow\; \alpha z_1 + (1-\alpha) z_2 \in \mathcal{S}. \qquad \text{(C.2)}$$

Figure C.1 illustrates this definition by showing two convex sets and one non-convex set. In the latter case, a segment is drawn between two points inside the set and it is seen that some of the points on the segment lie outside the set.

Convex Sets

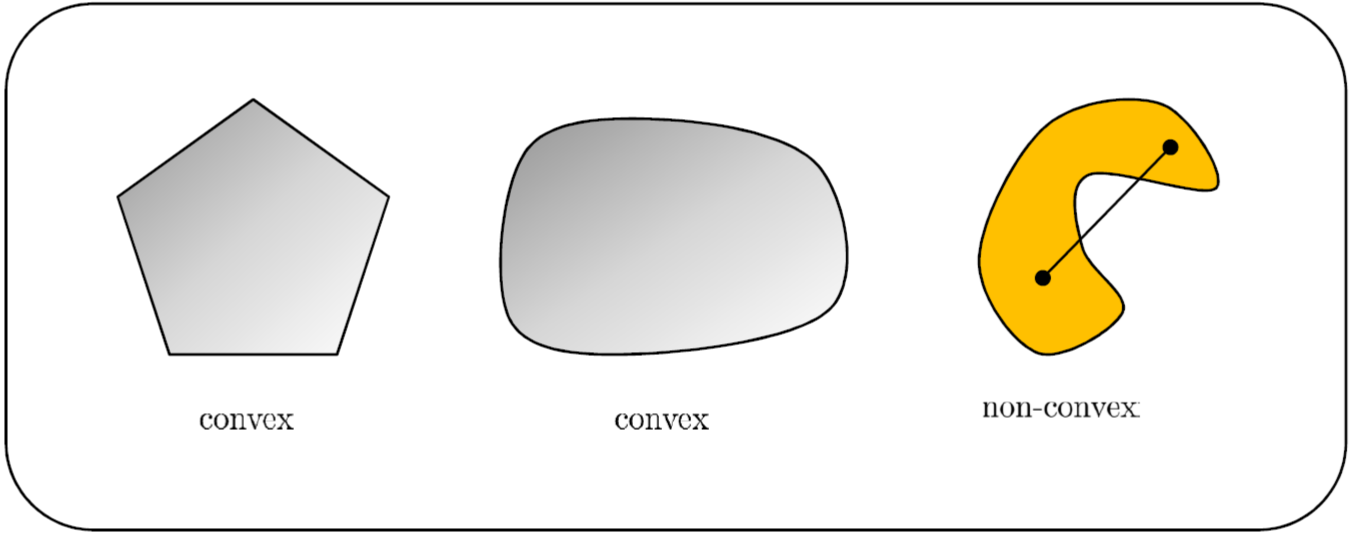

Figure C.1: The two sets on the left are examples of convex sets, while the set on the right is a non-convex set.

Convex Functions

The function $g(z)$ is said to be convex if its domain, written as $dom(g)$, is a convex set and if for any points $z_1, z_2 \in dom(g)$ and for any $0 \leq \alpha \leq 1$, it holds that

$$g(\alpha z_1 + (1-\alpha)z_2) \;\leq\; \alpha g(z_1) \;+\; (1-\alpha)g(z_2) \qquad \text{(C.3)}$$

In other words, all points belonging to the line segment connecting $g(z_1)$ to $g(z_2)$ lie on or above the graph of $g(z)$ — see the plot on the left side of Figure C.2.

Convex Functions

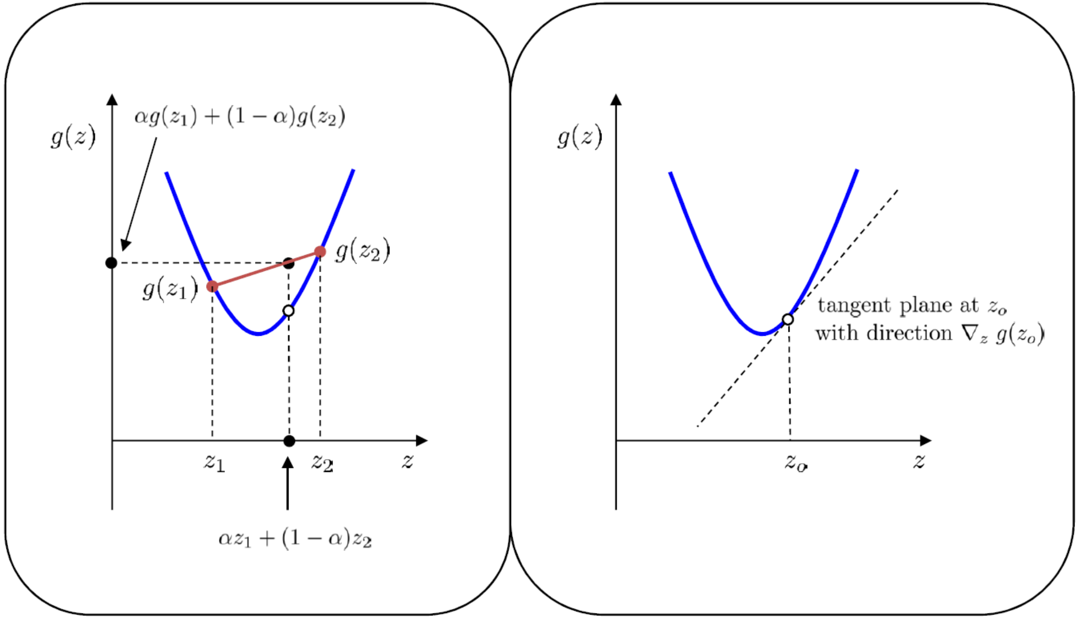

Figure C.2: Two equivalent characterizations of convexity for *differentiable* functions $g(z)$ as defined by (C.3) and (C.4).

Convex Functions

An equivalent characterization of convexity is that for any z_o and z:

$$g(z) \;\geq\; g(z_o) \;+\; [\nabla_z\, g(z_o)]\,(z - z_o) \qquad \text{(C.4)}$$

in terms of the inner product between the gradient vector at z_o and the vector difference $(z - z_o)$. This condition means that the tangent plane at z_o lies beneath the graph of the function — see the plot on the right side of Figure C.2.

Convex Functions

A useful property of every convex function is that, when a minimum exists, it can only be a global minimum; there can be multiple global minima but no local minima. That is, any stationary point at which the gradient vector of $g(z)$ is annihilated can only correspond to a global minimum of the function; the function cannot have local maxima, minima, or saddle points. A second useful property of convex functions, and which follows from (C.4), is that for any z_1 and z_2:

$$g(z) \text{ convex } \Longrightarrow \left[\nabla_z\, g(z_2) - \nabla_z\, g(z_1)\right](z_2 - z_1) \geq 0 \qquad \text{(C.5)}$$

Convex Functions

in terms of the inner product between two differences: the difference in the gradient vectors and the difference in the vectors themselves. The above result means that these difference vectors are aligned (i.e., have a nonnegative inner product). Result (C.5) follows by using (C.4) to write

$$g(z_2) \;\geq\; g(z_1) + [\nabla_z\, g(z_1)]\,(z_2 - z_1) \tag{C.6}$$

$$g(z_1) \;\geq\; g(z_2) + [\nabla_z\, g(z_2)]\,(z_1 - z_2) \tag{C.7}$$

Convex Functions

so that upon substitution of the second inequality into the right-hand side of the first inequality we obtain

$$g(z_2) \;\geq\; g(z_2) \;+\; [\nabla_z\, g(z_2)]\,(z_1 - z_2) \;+\; [\nabla_z\, g(z_1)]\,(z_2 - z_1) \qquad \text{(C.8)}$$

from which we obtain (C.5).

Example #C.1

Example C.1 (Convexity and sub-gradients). Property (C.4) is stated in terms of the gradient vector of $g(z)$ evaluated at location z_o. This gradient vector exists because we assumed the function $g(z)$ to be differentiable. There exist, however, cases where the function $g(z)$ need not be differentiable at all points. For example, for scalar arguments z, the function $g(z) = |z|$ is convex but is not differentiable at $z = 0$. For such non-differentiable convex functions, the characterization (C.4) can be replaced by the statement that the function $g(z)$ is convex if, and only if, for every z_o, a row vector $y \in \partial g(z_o)$ can be found such that

$$g(z) \geq g(z_o) + y(z - z_o) \tag{C.9}$$

Example #C.1

in terms of the inner product between y and the vector difference $(z - z_o)$. The vector y is called a sub-gradient and the notation $\partial g(z_o)$ denotes the set of all possible sub-gradients, also called the sub-differential of $g(z)$ at $z = z_o$; this situation is illustrated in Figure C.3. When $g(z)$ is differentiable at z_o, then there is a unique sub-gradient vector and it coincides with $\nabla_z\, g(z_o)$. In that case, statement (C.9) reduces to (C.4). We continue our presentation by focusing on differentiable functions $g(z)$.

$\blacksquare$

Example #C.1

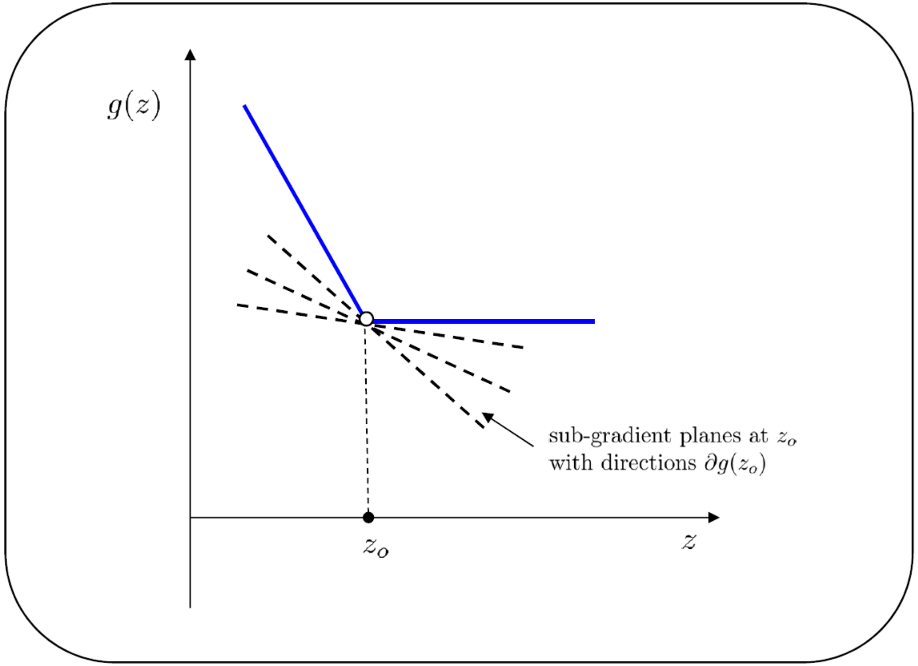

Figure C.3: A non-differentiable convex function with a multitude of sub-gradient directions at the point of non-differentiability.

Example #C.2

Example C.2 (Some useful operations that preserve convexity). It is straightforward to verify from the definition (C.3) that the following operations preserve convexity:

(1) if $g(z)$ is convex then $h(z) = g(Az + b)$ is also convex for any constant matrix A and vector b. That is, affine transformations of z do not destroy convexity.

(2) If $g_1(z)$ and $g_2(z)$ are convex functions, then $h(z) = \max\{g_1(z), g_2(z)\}$ is convex. That is, pointwise maximization does not destroy convexity.

(3) If $g_1(z)$ and $g_2(z)$ are convex functions, then $h(z) = a_1 g_1(z) + a_2 g_2(z)$ is also convex for any nonnegative coefficients a_1 and a_2.

■

Strict Convexity

The function $g(z)$ is said to be *strictly* convex if the inequalities in (C.3) or (C.4) are replaced by *strict* inequalities. More specifically, for any $z_1 \neq z_2$ and $0 < \alpha < 1$, a strictly convex function should satisfy:

$$g(\alpha z_1 + (1-\alpha)z_2) \;<\; \alpha g(z_1) \;+\; (1-\alpha)g(z_2) \tag{C.10}$$

A useful property of every strictly convex function is that, when a minimum exists, then it is both *unique* and also the global minimum of the function. A second useful property replaces (C.5) by the following statement with a strict inequality for any $z_1 \neq z_2$:

$$g(z) \text{ strictly convex } \Longrightarrow [\nabla_z\, g(z_2) - \nabla_z\, g(z_1)]\,(z_2 - z_1) > 0 \tag{C.11}$$

Strong Convexity

The function $g(z)$ is said to be *strongly* convex (or, more specifically, ν−strongly convex) if it satisfies the following stronger condition for any $0 \leq \alpha \leq 1$:

$$g(\alpha z_1 + (1-\alpha)z_2) \;\leq\; \alpha g(z_1) \;+\; (1-\alpha)g(z_2) \;-\; \frac{\nu}{2}\alpha(1-\alpha)\|z_1 - z_2\|^2 \tag{C.12}$$

for some scalar $\nu > 0$, and where the notation $\|\cdot\|$ denotes the Euclidean norm of its vector argument; although strong convexity can also be defined relative to other vector norms, the Euclidean norm is sufficient for our purposes.

Strong Convexity

Comparing (C.12) with (C.10) we conclude that strong convexity implies strict convexity. Therefore, every strongly convex function has a unique global minimum as well. Nevertheless, strong convexity is a stronger condition than strict convexity so that functions exist that are strictly convex but not necessarily strongly convex. For example, for scalar arguments z, the function $g(z) = z^4$ is strictly convex but not strongly convex. On the other hand, the function $g(z) = z^2$ is strongly convex — see Figure C.4. In summary, it holds that:

$$\text{strong convexity} \implies \text{strict convexity} \implies \text{convexity} \qquad \text{(C.13)}$$

Strong Convexity

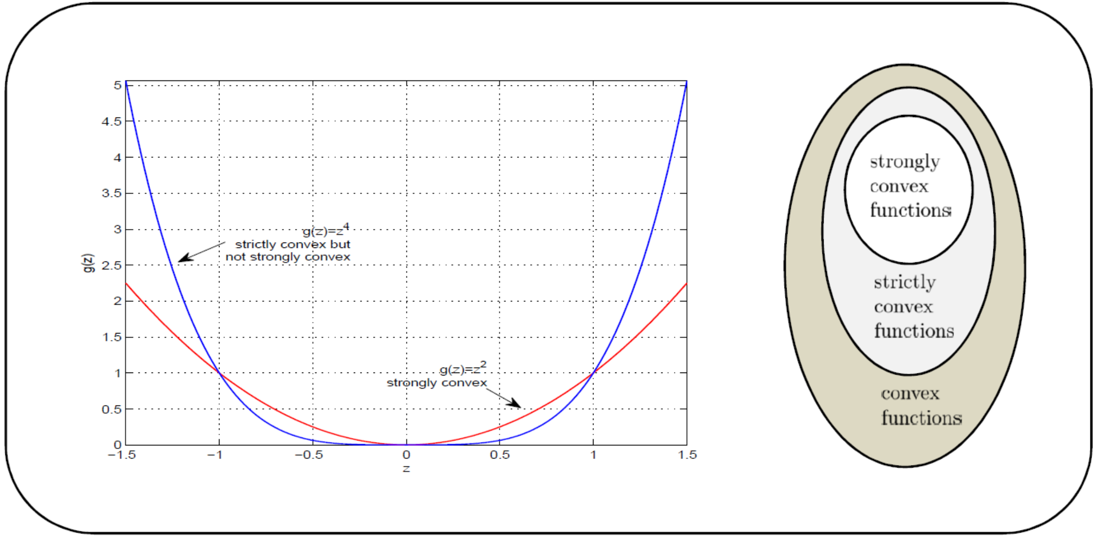

Figure C.4: The function $g(z) = z^4$ is strictly convex but not strongly convex, while the function $g(z) = z^2$ is strongly convex. Observe how $g(z) = z^4$ is more flat around its global minimizer and moves away from it more slowly than in the quadratic case.

Strong Convexity

A useful property of strongly convex functions is that they grow faster than a linear function in z since an equivalent characterization of strong convexity is that for any z_o and z:

$$g(z) \;\geq\; g(z_o) \;+\; [\nabla_z \, g(z_o)]\,(z - z_o) \;+\; \frac{\nu}{2}\|z - z_o\|^2 \qquad \text{(C.14)}$$

Strong Convexity

This means that the graph of $g(z)$ is strictly above the tangent plane at location z_o and moreover, for any z, the distance between the graph and the corresponding point on the tangent plane is at least as large as the quadratic term $\frac{\nu}{2}\|z - z_o\|^2$. In particular, if we specialize (C.14) to the case in which z_o is selected to correspond to the global minimizer of $g(z)$, i.e., as

$$z_o = z^o, \quad \text{where} \quad \nabla_z\, g(z^o) = 0 \tag{C.15}$$

Strong Convexity

then we conclude that every strongly convex function satisfies the following useful property for every z:

$$g(z) \; - \; g(z^o) \geq \frac{\nu}{2}\|z - z^o\|^2, \quad (z^o \text{ is global minimizer}) \qquad \text{(C.16)}$$

This property is illustrated in Figure C.5. Another useful property that follows from (C.14) is that for any z_1, z_2:

$$g(z) \text{ strongly convex} \Longrightarrow [\nabla_z \, g(z_2) - \nabla_z \, g(z_1)] \, (z_2 - z_1) \geq \nu \|z_2 - z_1\|^2 \qquad \text{(C.17)}$$

Strong Convexity

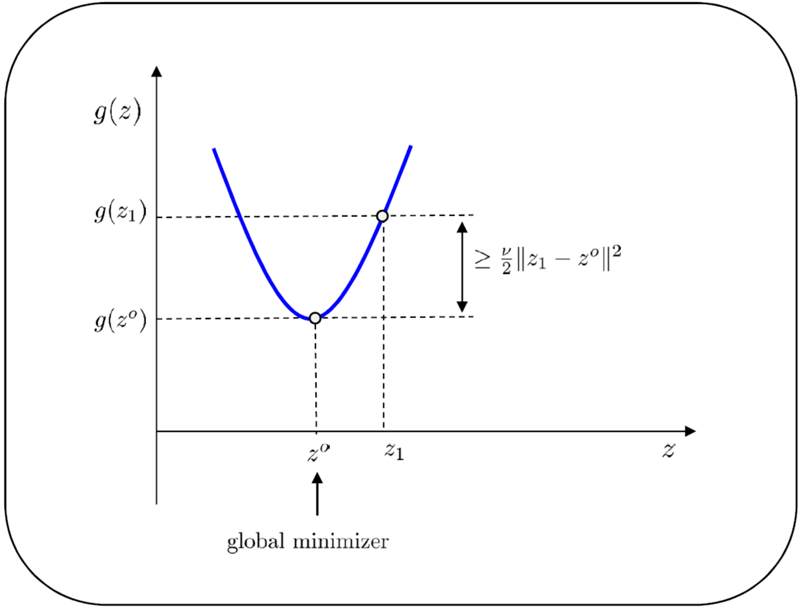

Figure C.5: For ν–strongly convex functions, the increment $g(z_1) - g(z^o)$ grows at least as fast as the quadratic term $\frac{\nu}{2}\|z_1 - z^o\|^2$, as indicated by (C.16) and where z^o is the global minimizer of $g(z)$.

Summary of Properties

Table C.1: Useful properties implied by the convexity, strict convexity, or strong convexity of a real-valued function $g(z) \in \mathbb{R}$ of a *real* argument $z \in \mathbb{R}^M$.

$$
\begin{aligned}
g(z) \text{ convex} \Longrightarrow &\quad [\nabla_z\, g(z_2) - \nabla_z\, g(z_1)]\,(z_2 - z_1) \geq 0 \\
g(z) \text{ strictly convex} \Longrightarrow &\quad [\nabla_z\, g(z_2) - \nabla_z\, g(z_1)]\,(z_2 - z_1) > 0 \\
g(z)\ \nu\text{-strongly convex} \Longrightarrow &\quad [\nabla_z\, g(z_2) - \nabla_z\, g(z_1)]\,(z_2 - z_1) \geq \nu \|z_2 - z_1\|^2
\end{aligned}
$$

Hessian Matrix Conditions

We indicated earlier that it is sufficient for our treatment to assume that the real-valued function $g(z)$ is differentiable whenever necessary. In particular, when it is twice continuously differentiable, then the properties of convexity, strict convexity, and strong convexity can be inferred from the Hessian matrix of $g(z)$ as follows (see, e.g., [178, 191]):

$$\begin{cases} \text{(a)} \;\; \nabla_z^2\, g(z) \geq 0 \text{ for all } z & \Longleftrightarrow \;\; g(z) \text{ is convex.} \\ \text{(b)} \;\; \nabla_z^2\, g(z) > 0 \text{ for all } z & \Longrightarrow \;\; g(z) \text{ is strictly convex.} \\ \text{(c)} \;\; \nabla_z^2\, g(z) \geq \nu I_M > 0 \text{ for all } z & \Longleftrightarrow \;\; g(z) \text{ is } \nu\text{--strongly convex.} \end{cases} \tag{C.18}$$

Hessian Matrix Conditions

Since $g(z)$ is real-valued and z is also real-valued in this section, then the Hessian matrix in this case is $M \times M$ and given by the expression shown in the first row of Table B.1 and by equation (B.29), namely,

$$\nabla_z^2\, g(z) \;\stackrel{\Delta}{=}\; \nabla_{z^\mathsf{T}} \left[\nabla_z\, g(z)\right] \qquad \text{(C.19)}$$

Hessian Matrix Conditions

Observe from (C.18) that the positive definiteness of the Hessian matrix is only a sufficient condition for strict convexity (for example, the function $g(z) = z^4$ is strictly convex even though its second-order derivative is not strictly positive for all z). One of the main advantages of working with strongly convex functions is that their Hessian matrices are sufficiently bounded away from zero.

Example #C.3

Example C.3 (Strongly-convex functions). The following is a list of useful strongly-convex functions that appear in applications involving adaptation, learning, and estimation:

(1) Consider the quadratic function

$$g(z) \;=\; \kappa + a^{\mathsf{T}} z + z^{\mathsf{T}} a + z^{\mathsf{T}} C z \qquad \text{(C.20)}$$

with a symmetric positive-definite matrix C. The Hessian matrix is $\nabla_z^2\, g(z) = 2C$, which is sufficiently bounded away from zero for all z since

$$\nabla_z^2\, g(z) \;\geq\; 2\lambda_{\min}(C)\, I_M > 0 \qquad \text{(C.21)}$$

in terms of the smallest eigenvalue of C. Therefore, such quadratic functions are strongly convex.

Example #C.3

(2) The regularized logistic (or log-)loss function

$$g(z) \;=\; \ln\left(1+e^{-\gamma h^{\mathsf T} z}\right) \;+\; \frac{\rho}{2}\|z\|^2 \qquad \text{(C.22)}$$

with a scalar γ, column vector h, and $\rho > 0$ is also strongly convex. This is because the Hessian matrix is given by

$$\nabla_z^2\, g(z) \;=\; \rho\, I_M \;+\; hh^{\mathsf T}\left(\frac{e^{-\gamma h^{\mathsf T} z}}{(1+e^{-\gamma h^{\mathsf T} z})^2}\right) \;\geq\; \rho\, I_M \;>\; 0 \qquad \text{(C.23)}$$

Example #C.3

(3) The regularized hinge loss function

$$g(z) \;=\; \max\left\{0, 1-\gamma h^{\mathsf{T}} z\right\} \;+\; \frac{\rho}{2}\|z\|^2 \tag{C.24}$$

with a scalar γ, column vector h, and $\rho > 0$ is also strongly convex, although non-differentiable. This result can be verified by noting that the function $\max\left\{0, 1-\gamma h^{\mathsf{T}} z\right\}$ is convex in z while the regularization term $\frac{\rho}{2}\|z\|^2$ is $\rho-$strongly convex in z.

$\blacksquare$

Complex Domain

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.

Convexity in Complex Domain

We now extend the previous definitions and results to the case in which $z \in \mathbb{C}^M$ is complex-valued, while $g(z) \in \mathbb{R}$ continues to be real-valued. One way to extend the concepts of convexity, strict convexity, and strong convexity to the case of complex arguments is to view $g(z)$ as a function of the extended real variable $v = \text{col}\{x, y\} \in \mathbb{R}^{2M}$, i.e., to work with $g(v)$ instead of $g(z)$, where v is defined in terms of the real and imaginary parts of z, namely, $z = x + jy$. Observe in particular that the complex variables z and z^* can be recovered from knowledge of v as follows:

Convexity in Complex Domain

$$\underbrace{\begin{bmatrix} I_M & jI_M \\ I_M & -jI_M \end{bmatrix}}_{=D}\underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{=v} = \begin{bmatrix} z \\ (z^*)^{\mathsf{T}} \end{bmatrix} \tag{C.25}$$

where the matrix D was introduced earlier in (B.27).

Convex Functions

The function $g(z)$ is said to be convex in z if the corresponding function $g(v)$ is convex in v, i.e., if $dom(g(v))$ is a convex set and for any $v_1, v_2 \in dom(g(v))$ and any $0 \leq \alpha \leq 1$, it holds that:

$$g(\alpha v_1 + (1-\alpha)v_2) \;\leq\; \alpha g(v_1) \;+\; (1-\alpha)g(v_2) \tag{C.26}$$

Since $g(z)$ is real-valued, the above condition can be restated in terms of the original complex variables $z_1, z_2 \in \mathbb{C}^M$ as follows:

$$g(\alpha z_1 + (1-\alpha)z_2) \;\leq\; \alpha g(z_1) \;+\; (1-\alpha)g(z_2) \tag{C.27}$$

Convex Functions

An equivalent characterization of the convexity condition (C.26) is that for any v_o,

$$g(v) \;\geq\; g(v_o) \;+\; [\nabla_v\, g(v_o)]\,(v - v_o) \qquad \text{(C.28)}$$

This condition can again be restated in terms of the original complex variables $\{z, z_o\}$. To do so, we first need to find the relation between the gradient vector $\nabla_v\, g(v)$ evaluated in the $v-$domain and the gradient vector $\nabla_z\, g(z)$ evaluated in the $z-$domain. Thus, recall that v is a column vector obtained by stacking x and y on top of each other. Therefore, by referring to definition (A.26), we have that

$$\nabla_v\, g(v) \;=\; \left[\begin{array}{cc} \nabla_x\, g(x,y) & \nabla_y\, g(x,y) \end{array}\right] \qquad \text{(C.29)}$$

Convex Functions

Multiplying from the right by the matrix D^* from (B.27) we obtain

$$\nabla_v\, g(v) \cdot \frac{1}{2} D^* = \frac{1}{2} \left[\; \nabla_x\, g(x,y) \quad \nabla_y\, g(x,y) \;\right] \left[\begin{array}{cc} I_M & I_M \\ -jI_M & jI_M \end{array} \right] \qquad \text{(C.30)}$$

Now consider the following complex gradient vectors, which correspond to the extension of the earlier definition (A.9) to the vector case for real-valued functions $g(z)$:

$$\left\{ \begin{array}{lcl} \nabla_z\, g(z) & \stackrel{\Delta}{=} & \frac{1}{2}\left[\nabla_x\, g(x,y) \;-\; j\nabla_y\, g(x,y)\right] \\ \nabla_{z^*}\, g(z) & \stackrel{\Delta}{=} & \frac{1}{2}\left[\nabla_{x^\mathsf{T}}\, g(x,y) \;+\; j\nabla_{y^\mathsf{T}}\, g(x,y)\right] \end{array} \right. \qquad \text{(C.31)}$$

Convex Functions

Substituting into the right-hand side of (C.30) we conclude that

$$\frac{1}{2}\left[\nabla_v\, g(v)\right] D^* \;=\; \left[\;\nabla_z\, g(z) \qquad \left(\nabla_{z^*}\, g(z)\right)^{\mathsf{T}}\;\right] \tag{C.32}$$

which is the desired relation between the gradient vectors $\nabla_v\, g(v)$ and $\nabla_z\, g(z)$. Using (C.25) and (C.32), and noting that $g(z) = g(v)$, we can now rewrite (C.28) in terms of the original complex variables $\{z, z_o\}$ as follows:

$$g(z) \;\geq\; g(z_o) \;+\; 2\mathrm{Re}\left\{\,\left[\nabla_z\, g(z_o)\right](z - z_o)\,\right\} \tag{C.33}$$

Convex Functions

in terms of the real part of the inner product that appears on the right-hand side. A useful property that follows from (C.33) is that for any z_1 and z_2 :

$$g(z) \text{ convex} \Longrightarrow \text{Re}\left\{ \left[\nabla_z\, g(z_2) - \nabla_z\, g(z_1)\right](z_2 - z_1) \right\} \geq 0 \qquad \text{(C.34)}$$

Strict Convexity

The function $g(z)$ is said to be *strictly* convex if the inequalities in (C.27) or (C.33) are replaced by strict inequalities. For example, for any $z_1 \neq z_2$ and $0 < \alpha < 1$, a strictly convex function $g(z)$ should satisfy:

$$g(\alpha z_1 + (1-\alpha) z_2) \; < \; \alpha g(z_1) \; + \; (1-\alpha) g(z_2) \tag{C.35}$$

Again, a useful property of every strictly convex function is that, when a minimum exists, then it is both unique and the global minimum of the function. Another useful property is that for any $z_1 \neq z_2$:

$$g(z) \text{ strictly convex} \Longrightarrow \text{Re}\left\{ \left[\nabla_z \, g(z_2) - \nabla_z \, g(z_1) \right] (z_2 - z_1) \right\} > 0 \tag{C.36}$$

Strong Convexity

The function $g(z)$ is said to be *strongly* convex (or, more specifically, $\nu-$strongly convex) in z if $g(v)$ is $\nu-$strongly convex in v, i.e., if $g(v)$ satisfies the following condition for any $0 \leq \alpha \leq 1$:

$$g(\alpha v_1 + (1-\alpha)v_2) \;\leq\; \alpha g(v_1) \;+\; (1-\alpha)g(v_2) \;-\; \frac{\nu}{2}\alpha(1-\alpha)\|v_1 - v_2\|^2 \tag{C.37}$$

for some $\nu > 0$. Using the fact that

$$\|v_1 - v_2\|^2 \;=\; \|z_1 - z_2\|^2 \tag{C.38}$$

Strong Convexity

the above condition can be restated in terms of the original complex variables as follows:

$$g(\alpha z_1+(1-\alpha)z_2) \leq \alpha g(z_1)+(1-\alpha)g(z_2)-\frac{\nu}{2}\alpha(1-\alpha)\|z_1-z_2\|^2 \quad \text{(C.39)}$$

An equivalent characterization of strong convexity is that for any z_o,

$$g(z) \;\geq\; g(z_o) \;+\; 2\mathrm{Re}\left\{\left[\nabla_z\, g(z_o)\right](z-z_o)\right\} \;+\; \frac{\nu}{2}\|z-z_o\|^2 \qquad \text{(C.40)}$$

Strong Convexity

In particular, if we select z_o to correspond to the global minimizer of $g(z)$, i.e.,

$$z_o = z^o \qquad \text{where} \qquad \nabla_z\, g(z^o) = 0 \tag{C.41}$$

then strongly convex functions satisfy the following useful property:

$$g(z) \;-\; g(z^o) \geq \frac{\nu}{2}\|z - z^o\|^2, \qquad (z^o \text{ is global minimizer}) \tag{C.42}$$

Strong Convexity

Another useful property that follows from (C.40) is that for any z_1, z_2:

$$g(z) \text{ strongly convex} \Longrightarrow$$

$$\text{Re}\left\{\left[\nabla_z\, g(z_2) - \nabla_z\, g(z_1)\right](z_2 - z_1)\right\} \geq \frac{\nu}{2}\|z_2 - z_1\|^2 \tag{C.43}$$

Summary of Properties

Table C.2: Useful properties implied by the convexity, strict convexity, or strong convexity of a real-valued function $g(z) \in \mathbb{R}$ of a *complex* argument $z \in \mathbb{C}^M$.

$$g(z) \text{ convex} \Longrightarrow \quad \text{Re}\{\,[\nabla_z\, g(z_2) - \nabla_z\, g(z_1)]\,(z_2 - z_1)\,\} \geq 0$$

$$g(z) \text{ strictly convex} \Longrightarrow \quad \text{Re}\{[\nabla_z\, g(z_2) - \nabla_z\, g(z_1)]\,(z_2 - z_1)\} > 0$$

$$g(z)\ \nu\text{−strongly convex} \Longrightarrow \quad \text{Re}\{[\nabla_z\, g(z_2) - \nabla_z\, g(z_1)]\,(z_2 - z_1)\} \geq \frac{\nu}{2}\|z_2 - z_1\|^2$$

Hessian Matrix Conditions

Since $g(z)$ is real-valued and z is now complex-valued, then the Hessian matrix of $g(z)$ is $2M \times 2M$ and given by the expression shown in the last row of Table B.1 — see (B.29). As before, the properties of convexity, strict convexity, and strong convexity can be inferred from the Hessian matrix of $g(z)$ as follows:

$$\left\{ \begin{array}{llll} \text{(a)} & \nabla_z^2\, g(z) \geq 0 \text{ for all } z & \Longleftrightarrow & g(z) \text{ is convex.} \\ \text{(b)} & \nabla_z^2\, g(z) > 0 \text{ for all } z & \Longrightarrow & g(z) \text{ is strictly convex.} \\ \text{(c)} & \nabla_z^2\, g(z) \geq \frac{\nu}{2} I_{2M} > 0 \text{ for all } z & \Longleftrightarrow & g(z) \text{ is strongly convex.} \end{array} \right. \tag{C.44}$$

Hessian Matrix Conditions

Observe again that the positive definiteness of the Hessian matrix is only a sufficient condition for strict convexity. Moreover, the condition in part (c), with a factor of $\frac{1}{2}$ multiplying ν, follows from the following sequence of arguments:

$$
\begin{aligned}
g(z) \text{ is } \nu\text{−strongly convex} \quad &\Longleftrightarrow \quad g(v) \text{ is } \nu\text{−strongly convex} \\
&\overset{(C.18)}{\Longleftrightarrow} \quad H(v) \geq \nu I_{2M} > 0, \quad \text{for all } v \\
&\Longleftrightarrow \quad \frac{1}{4} D H(v) D^* \geq \frac{\nu}{4} D D^* \overset{(B.28)}{=} \frac{\nu}{2} I > 0 \\
&\overset{(B.26)}{\Longleftrightarrow} \quad H_c(u) \geq \frac{\nu}{2} I_{2M} > 0 \qquad \text{(C.45)}
\end{aligned}
$$

Example #C.4

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

$$g(z) = \kappa + a^* z + z^* a + z^* C z \tag{C.46}$$

with a Hermitian positive-definite matrix $C > 0$. The complex Hessian matrix is given by

$$H_c(u) \;=\; \begin{bmatrix} C & 0 \\ 0 & C^{\mathsf{T}} \end{bmatrix} \tag{C.47}$$

which is sufficiently bounded away from zero from below since

$$H_c(u) \;\geq\; \lambda_{\min}(C)\, I_{2M} > 0 \tag{C.48}$$

Therefore, such quadratic functions are strongly convex. ■

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.