CISC 371 Class 18

Convex Functions, Convex Sets, and Level Sets

Texts: [1] pp. 51–60; [2] pp. 117–132; [3] pp. 67–87

Main Concepts:

- *Convex functions: affine and quadratic*
- *Convex functions: positive scaling and addition*
- *Convex function: above tangent plane*
- *Level set:* $\mathbb{S}_S(f, l)$
- *Convex set: contains linear interpolant*
- *Property: level set of a convex function is a convex set*

Sample Problem, Optimization: What is the difference between a function that steadily decreases towards a minimum, and a convex function?

Convexity is a property of many mathematical objects, including functions and sets. Entire textbooks – such as one of the texts recommended as a reference in this course – are devoted to optimization of convex functions. Some of the major current work in machine learning and data analytics requires optimization where the search vector is *constrained*.

To understand how we can solve a constrained optimization problem, it will be helpful for us to understand the concept of convexity in a deeper sense than we have so far in this course. Recall that, in Definition 3.6, we defined a convex function. Replacing the inequality in the consequent term of the implication in Definition 3.6 provides a definition of a *strictly convex* function.

A basic distinction we need to make is between a function that is *monotonic* and a function that is *convex*. A simple example is the Gaussian function; we will negate it, or turn it "upside down", so that it fits our definition. The equation for the negation of the probability density function, with zero mean and unit variance, is

$$f(t) = \frac{-1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \tag{18.1}$$

This inverted Gaussian is plotted in Figure 18.1. This function monotonically decreases from 0 to $-1/\sqrt{2\pi}$ from $t = -\infty$ or from $t = +\infty$. The function is *not* convex because the line segment from $u = -0.5$ to $v = +3$ intersects the graph of the function: this function is not always "below" a chord that is drawn between two points.

A convex function that has a scalar argument can be defined as a function that is always "above" a *chord*, which is a line segment that joins two points.

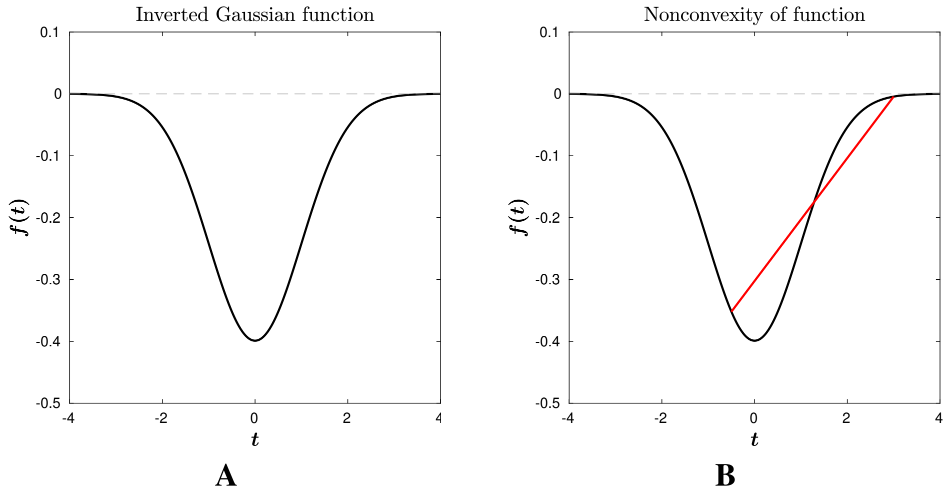


Figure 18.1: An inverted Gaussian is a function that monotonically decreases towards a minimum and that is not convex. (A) The function $f(t)$ in Equation 18.1. (B) A chord between two points that intersects the graph of the function.

Convexity is crucial to understanding many of the concepts in optimization, particularly optimization that is mathematical constrained. For a function with a vector argument, the definition is a direct extension of Definition 3.6.

Definition: convex function $f(\vec{w})$

For any $\vec{u} \in \mathbb{R}^n$, any $\vec{b} \in \mathbb{R}^n$, any $\theta \in \mathbb{R}_+$, and any $f : \mathbb{R}^n \to \mathbb{R}$, that function f is a *convex function* is defined as

$$(0 \leq \theta \leq 1) \to \left(f((1-\theta)\vec{u} + \theta\vec{b})\right) \leq \left((1-\theta)f(\vec{u}) + \theta f(\vec{b})\right) \tag{18.2}$$

Observations: The line segment that "connects" the points $\vec{u}$ and $\vec{b}$ is in the vector space $\mathbb{R}^n$. The result of evaluating the function at a point in its domain is a scalar, so the inequality constraint is a scalar constraint.

The definition of a *strictly convex function* is also a straightforward extension of the scalar definition. In this course, we will not often need the definition of strictly convex. Definition 18.14 is in the extra notes for this class as a matter of completeness.

A second definition that is in the extra notes is Definition 18.15, which defines convexity for a vector function with a vector argument. In plain English, such a function is convex is defined as having each "component" function being convex.

18.1 Convexity of Affine and Quadratic Functions

Two kinds of functions will occur so often in this course that we will prove that they are convex. The first kind is an *affine* function, which is the product of a matrix and a vector to which a second vector is added. An affine function has the entire vector space $\mathbb{R}^n$ as its domain.

Theorem: An affine map, $f(\vec{w}) = \underline{m}\vec{w} + c$, is convex.

Proof: See the extra notes for this class.

A straightforward extension of Theorem 18.16 is that a general linear map is convex; this is stated and proved in the extra notes for this class.

The second kind of convex function that we will use is a *quadratic form* of a matrix. We can assume, without loss of generalization, that such a real matrix K is a symmetric matrix.

Theorem: A quadratic form of a positive definite matrix, $f(\vec{w}) = \vec{w}^T K \vec{w}$, is convex.

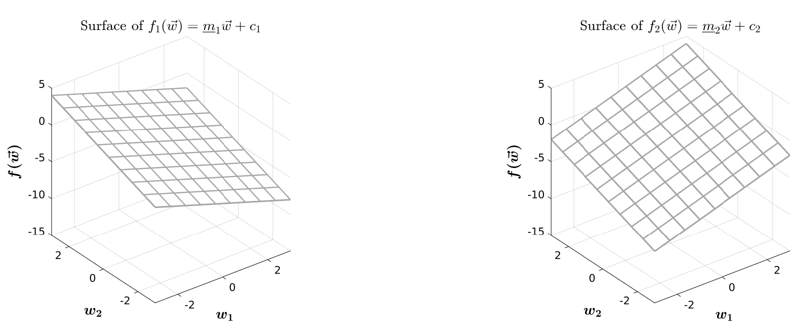
Proof: See the extra notes for this class.

We can visualize a convex function, that has a 2D vector argument, using ordinary plotting software such as `MATLAB`. Two affine functions that we might consider are

$$f_1(\vec{w}) = \underline{m}_1\vec{w} + c_1 \quad \text{where} \quad \underline{m}_1 = \begin{bmatrix} -1 & 1 \end{bmatrix} \quad \text{and} \quad c_1 = -2 \tag{18.3}$$

$$f_2(\vec{w}) = \underline{m}_2\vec{w} + c_2 \quad \text{where} \quad \underline{m}_2 = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad \text{and} \quad c_2 = +2 \tag{18.4}$$

These functions are plotted in Figure 18.2.



A **B**

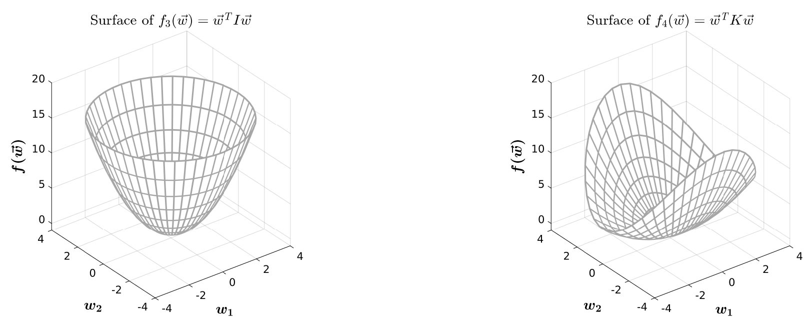
Figure 18.2: Affine functions, each with a 2D vector argument; these plot as planes in 3D. (A) The function $f_1(\vec{w})$ defined in Equation 18.3. (B) The function $f_2(\vec{w})$ defined in Equation 18.3.

We can also visualize a quadratic function that has a 2D vector argument. Two examples are

$$f_3(\vec{w}) \quad = \quad \vec{w}^T\vec{w} = \vec{w}^T I \vec{w} \tag{18.5}$$

$$f_4(\vec{w}) \quad = \quad \vec{w}^T K \vec{w} \quad \text{where} \quad K = \begin{bmatrix} 1/4 & 0 \\ 0 & 1 \end{bmatrix} \tag{18.6}$$

These functions are plotted in Figure 18.3.



A **B**

Figure 18.3: Quadratic functions, each with a 2D vector argument; these plot as paraboloids in 3D. (A) The function $f_3(\vec{w})$ defined in Equation 18.5 is a surface of revolution. (B) The function $f_4(\vec{w})$ defined in Equation 18.6 is not symmetric about the vertical axis.

18.2 Operations That Preserve Convexity of Functions

Of the many operations that preserve the convexity of functions, we will mainly use three. The first two are simple, have an intuitive appeal, and are easy to prove.

Recall, from basic mathematics, two properties of inequalities. The first is that an inequality is unchanged when both sides are multiplied by the same nonnegative number. Suppose that $l \in \mathbb{R}_+$, which is supposing that l is a real number and that $0 \le l$. for any real numbers a and b, a basic property is

$$(a \le b) \rightarrow (la \le lb) \tag{18.7}$$

Property 18.7 can be used to prove a basic convexity result.

Theorem: If f is a convex function, then for any nonnegative $l \in \mathbb{R}_+$, lf is a convex function.

Proof: See the extra notes for this class.

Next, recall that the sum of two inequalities preserves inequality. Suppose that for any real numbers a, b, c, and d, we have $a \le b$ and $c \le d$. Then it follows that $(a + c) \le (b + d)$. More formally, the property is

$$((a \leq b) \wedge (c \leq d)) \rightarrow ((a + c) \leq (b + d)) \tag{18.8}$$

Property 18.8 can be used to prove another basic convexity result.

Theorem: If f_1 is a convex function and f_2 is a convex function, then $(f_1 + f_2)$ is a convex function.

Proof: See the extra notes for this class.

Observation: These extend easily to nonnegative weighted sums of convex functions. Figure 18.4 shows the simple sum of the function $f_1(\vec{w})$ and $f_2(\vec{w})$ that we previously defined.

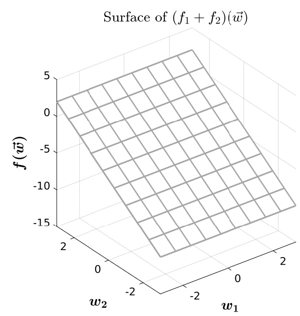


Figure 18.4: The sum of linear functions, $f_1(\vec{w})$ and $f_2(\vec{w})$, is a convex function. The plot of $(f_1 + f_2)(\vec{w})$ is a 2D plane in 3D.

Multiplication by a nonnegative number has intuitive appeal. Summation is, to some students, a less intuitive result.

Example: Because a quadratic form is convex, and an affine transformation is convex, the sum of these is convex. For a symmetric matrix $K \succ 0$ and a vector $\vec{q}$, the function

$$f(\vec{w}) = \frac{1}{2}\vec{w}^T K \vec{w} + \vec{q}^T \vec{w} + c$$

is a convex function. Specifically, for a "center" vector $\vec{w}_0$, the quadratic form is a convex function:

$$f(\vec{w}) = \frac{1}{2}[\vec{w} - \vec{w}_0]^T K [\vec{w} - \vec{w}_0]$$

A more complicated result that we will use in this course is variously called convexity for an affine map, or convexity for a change of variables. A linear map $\vec{v} = M\vec{u}$ can be thought of as a change from the variables in the vector $\vec{u}$ to the variables in the vector $\vec{v}$. In the context of linear maps, we usually think of the matrix M as being square and full rank.

Here, a matrix M can be any $m \times n$ real matrix. We can also have a vector $\vec{c} \in \mathbb{R}^m$. An affine map, or change of variables, is the transformation

$$\vec{v} = M\vec{u} + \vec{c} \tag{18.9}$$

Consider a convex function $f : \mathbb{R}^m \to \mathbb{R}$, which produces a scalar value for a vector $\vec{v} \in \mathbb{R}^m$. If we replace $\vec{v}$ by an affine map of $\vec{u}$, the composed function is a convex function.

Theorem: If f is a convex function then, for the affine map $M\vec{u} + \vec{c}$,

$g(\vec{u}) = f(M\vec{u} + \vec{c})$ is a convex function.

Proof: See the extra notes for this class.

Observation: Convexity for an affine change of variables, plus convexity of a norm, can be used to prove convexity of a quadratic form of a symmetric positive definite matrix. The Cholesky decomposition may be useful in such a proof.

We can use the Taylor series to write a function, then truncate the series after the first derivative. Because the sum of the truncated terms is nonnegative, this gives us a way to find a lower bound of a locally convex function. Consider a convex open set $\mathbb{V} \subseteq \mathbb{R}^n$, and a well behaved function $f : \mathbb{V} \to \mathbb{R}$ that is convex for any vector $\vec{w} \in \mathbb{V}$. The first few terms of the Taylor series are straightforward extensions of the case of a function with a scalar argument. From the assumption that f is convex over the convex set $\mathbb{V}$, the vector version of the Mean Value Theorem applies. For some point $\vec{\xi} \in \mathbb{V}$, the value of the function f at the point $\vec{w} \in \mathbb{V}$ can be expanded from a point $\vec{w}_0 \in \mathbb{V}$ as

$$f(\vec{w}) = f(\vec{w}_0) + \underline{\nabla} f(\vec{w}_0)(\vec{w} - \vec{w}_0) + \frac{1}{2}[\vec{w} - \vec{w}_0]^T \underline{\nabla}^2 f(\vec{\xi})[\vec{w} - \vec{w}_0] \tag{18.10}$$

Because we have assumed that f is convex over $\mathbb{V}$, we know that the Hessian matrix is symmetric and positive semidefinite when it is evaluated at any point in $\mathbb{V}$. The third term in Equation 18.10 is, consequently, nonnegative; we can thus deduce the inequality

$$f(\vec{w}) \geq f(\vec{w}_0) + \underline{\nabla} f(\vec{w}_0)(\vec{w} - \vec{w}_0) \tag{18.11}$$

In Equation 18.11, the gradient acts the same way that a *normal vector* acts: the scalar $\underline{\nabla} f(\vec{w}_0)(\vec{w} - \vec{w}_0)$ is zero if and only if $\vec{w}$ is in a plane that contains $\vec{w}_0$. We can conclude that a convex function is always "above" the *tangent hyperplane* that can be defined at any point in the domain set of the function. For the quadratic function $f_3(\vec{w})$ in Equation 18.11, this is illustrated in Figure 18.5

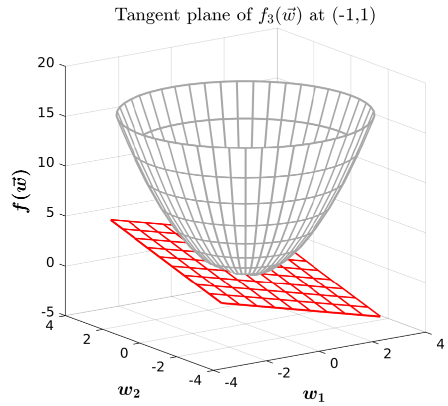


Figure 18.5: Because of the gradient inequality, the convex function $f_3(\vec{w})$ is always "above" its tangent plane at any point. The function is shown in black and the tangent plane at the point $\vec{w}_0 = (-1, 1)$ is shown in blue.

18.3 Level Sets

A concept that is closely associated to a level curve of a function is a *level set*. This is the set of all vectors that evaluate to a scalar that is less than, or equal to, a specified level.

Definition: level set of $f(\vec{w})$ at l is $\mathbb{S}_L(f, l)$

For any $\vec{u} \in \mathbb{R}^n$, any $l \in \mathbb{R}$, and any $f : \mathbb{R}^n \to \mathbb{R}$, the *level set* of f at l is defined as

$$\mathbb{S}_L(f, l) \stackrel{\text{def}}{=} \{\vec{u} : f(\vec{u}) \leq l\} \tag{18.12}$$

A level set "fills in" the region that is enclosed by a level curve. For examples of level sets, Figure 18.6 shows the surfaces and level sets of a linear function and a quadratic function.

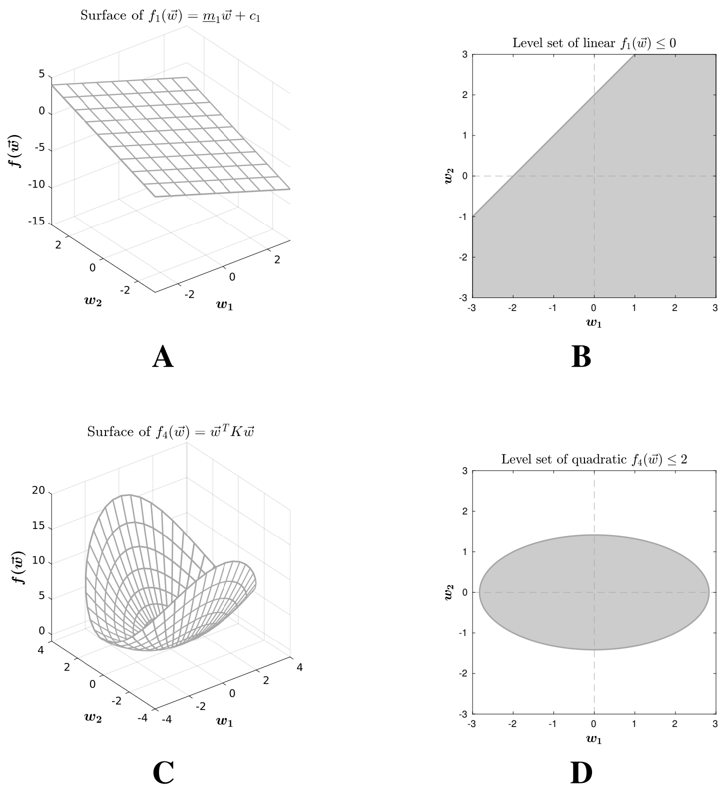


Figure 18.6: Surface plots and level sets of convex functions that have a 2D vector argument. (A) A surface plot of function $f_1(\vec{w})$ defined in Equation 18.3. (B) The level set of $f_1(\vec{w}) \leq 0$, written as $\mathbb{S}_L(f_1, 0)$, is shown as the shaded region; the blue arrow is the direction of the transpose of the gradient of the function. (C) A surface plot of function $f_4(\vec{w})$ defined in Equation 18.6. (D) The level set of $f_4(\vec{w}) \leq 2$, written as $\mathbb{S}_L(f_2, 2)$, is shown as the shaded region.

18.4 Convex Sets

Later in this course, we will use *convex sets* to describe how an optimization problem is constrained. The formal definition is conceptually similar to how we defined a convex function.

Definition: convex set $\mathbb{V}$

> For any $\mathbb{V} \subseteq \mathbb{R}^n$, any $\vec{u} \in \mathbb{V}$, any $\vec{v} \in \mathbb{V}$, and any $\theta \in \mathbb{R}_+$, that $\mathbb{V}$ is a *convex set* is defined as
>
> $$(0 \leq \theta \leq 1) \rightarrow (((1-\theta)\vec{u} + \theta\vec{v}) \in \mathbb{V}) \tag{18.13}$$

In plain English, this means that a set $\mathbb{V}$ is convex implies that the line segment between any two points in $\mathbb{V}$ lies entirely in $\mathbb{V}$.

In this course, we will mainly specify a convex set as either a level set of a convex function, or as derived from an operation on convex sets. There are two properties of convex sets that we will use in this course and that are proved in the extra notes for this class.

There are many other properties of convex sets that are needed for optimization of convex functions subject to convex constraints. An interested student should feel encouraged to explore the associated texts and other material on this extensive topic.

Theorem: the intersection of two convex sets is a convex set.

Proof: See the extra notes for this class.

Observations: This result is familiar from prerequisite material that includes Venn diagrams of up to three sets, which are typically shown as a convex sets. A consequence of Theorem 18.23, which we will use later in this course, is that the intersection of level sets of convex functions is a convex set. An example is the intersection of the level sets $f_1(\vec{w}) \leq 0$ and $f_2(\vec{w}) \leq 0$, shown in Figure 18.7.

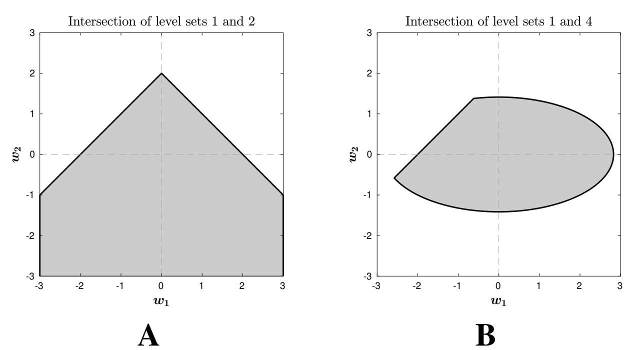


Figure 18.7: Level sets of convex functions that have a 2D vector argument, shown as shaded regions. (A) The intersection of the level sets $\mathbb{S}_L(f_1, 0)$ and $\mathbb{S}_L(f_2, 0)$ is a convex set. (B) The intersection of the level sets of $\mathbb{S}_L(f_1, 0)$ and $\mathbb{S}_L(f_4, 2)$ is a convex set.

Theorem: an affine transformation of a convex set is convex.

Proof: See the extra notes for this class.

Observations: This statement has a simple interpretation. Consider any convex set $\mathbb{V}$. The statement means that, if $\vec{w} \in \mathbb{V}$, then an affinely transformed $\vec{w}_A = M\vec{w} + \vec{c}$ is in some convex set $\mathbb{V}_A$. This implies that we can "add" a constant vector $\vec{c}$ to a convex set, which offsets each vector in $\mathbb{V}$ to a new vector. We can also multiply a vector in $\mathbb{V}$ by any real number – including a negative number – and the new set is also convex. Multiplication by the number -1 is the operation of reflecting the set around the origin, which maintains the convexity of the set.

18.5 Extra Notes on Convex Functions and Convex Sets

These extra notes are in three sub-sections: definitions related to convex functions; proofs that relevant functions are convex; and proofs related to convex sets.

18.5.1 Definitions for Convex Functions

Definition: strictly convex function $f(\vec{w})$

For any $\vec{u} \in \mathbb{R}^n$, any $\vec{v} \in \mathbb{R}^n$, any $\theta \in \mathbb{R}_{++}$, and any $f:\mathbb{R}^n \to \mathbb{R}$, that function f is a *strictly convex function* is defined as

$$((\vec{u} \neq \vec{v}) \wedge (0 < \theta < 1)) \to (f((1-\theta)\vec{u} + \theta\vec{v})) < ((1-\theta)f(\vec{u}) + \theta f(\vec{v}))) \quad (18.14)$$

Observation: Strict convexity of a function with a vector argument is a simple extension of the definition of a function with a scalar argument.

Definition: convex function $\vec{f}(\vec{w})$

For any $f:\mathbb{R}^n \to \mathbb{R}^m$ and any $i \in \mathbb{N}_{++}$, that function $\vec{f}$ is a *convex function* is defined as

$$(i \leq m) \to (f_i \text{ is convex}) \quad (18.15)$$

Observation: A vector function with a vector argument is convex if and only if each "component" function f_i is convex.

A convex function that we will occasionally encounter is a *norm* of a vector. Recall that one property of a norm is the triangle inequality, which is

$$\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$$

The triangle inequality holds for any norm, not only for the usual Euclidean norm.

18.5.2 Theorems for Convex Functions

Theorem: An affine map to scalars is convex.

For any $\underline{m} \in \mathbb{R}^n$ and any $c \in \mathbb{R}$, the affine map

$$f(\vec{w}) = \underline{m}\vec{w} + c \tag{18.16}$$

is convex.

Proof: Let $\vec{u} \in \mathbb{R}^n$, $\vec{v} \in \mathbb{R}^n$, and $\theta \in \mathbb{R}_+$ with $0 \leq \theta \leq 1$. Then

$$\begin{aligned}
f((1-\theta)\vec{u} + \theta\vec{v}) &= \underline{m}[(1-\theta)\vec{u} + \theta\vec{v}] + c \\
&= \underline{m}(1-\theta)\vec{u} + \underline{m}\theta\vec{v} + c \\
&= \underline{m}(1-\theta)\vec{u} + \underline{m}\theta\vec{v} + (1-\theta)c + \theta c \\
&= (1-\theta)\underline{m}\vec{u} + (1-\theta)c + \theta\underline{m}\vec{v} + \theta c \\
&= (1-\theta)f(\vec{u}) + \theta f(\vec{v})
\end{aligned}$$

so f is convex.

Observation: An affine map to a scalar value is convex and is not strictly convex. The latter property can be deduced because the linear combination of arguments equals the linear combination of the evaluations of the arguments.

Theorem: An affine map to a vector space is convex.

For any real $m \times n$ matrix M and any $\vec{c} \in \mathbb{R}^m$, the affine map

$$\vec{f}(\vec{w}) = M\vec{w} + \vec{c} \tag{18.17}$$

is convex.

Proof: Each "component" function f_i maps a vector $\vec{w} \in \mathbb{R}^n$ using row #i of M, which is $\underline{m}_i$, and entry c_i of $\vec{c}$ to a scalar value. Each f_i is convex by Theorem 18.17 so $\vec{f}$ is convex.

Observation: An affine map to a vector space is convex and is not strictly convex.

Theorem: A vector norm is convex.

For any norm $\|\cdot\|$ over a vector space $\mathbb{R}^n$, the function

$$f_N(\vec{w}) = \|\vec{w}\| \tag{18.18}$$

is convex.

Proof: Let $\vec{u} \in \mathbb{R}^n$, $\vec{v} \in \mathbb{R}^n$, and $\theta \in \mathbb{R}_+$ with $0 \leq \theta \leq 1$. Then

$$\begin{array}{rcl} f_N((1-\theta)\vec{u} + \theta\vec{v}) & = & \|(1-\theta)\vec{u} + \theta\vec{v}\| \\ & \leq & \|(1-\theta)[\vec{u}]\| + \|\theta\vec{v}\| \\ & \leq & (1-\theta)\|[\vec{u}]\| + \theta\|\vec{v}\| \\ & \leq & (1-\theta)f_N(\vec{u}) + \theta f_N(\vec{v}) \end{array}$$

Observation: Convexity of a norm can be useful for demonstrating convexity of other functions.

A prominent convex function is a *quadratic form* of a matrix. We can assume, without loss of generalization, that such a real matrix K is a symmetric matrix.

Theorem: A quadratic form of a positive definite matrix is convex.

For any $n \times n$ symmetric matrix $K \in (\mathbb{R}^n \times \mathbb{R}^n)$ for which $K \succ 0$, the quadratic form

$$f(\vec{w}) = \vec{w}^T K \vec{w} \tag{18.19}$$

is convex.

Proof: Let $\vec{u} \in \mathbb{R}^n$, $\vec{v} \in \mathbb{R}^n$, and $\theta \in \mathbb{R}_+$ with $0 \leq \theta \leq 1$. Observe that the inequality in the definition of a convex function has an equivalent inequality that holds because the same expression is added to each side of the inequality:

$$\begin{array}{rrcl} & (1-\theta)f(\vec{u}) + \theta f(\vec{v})) & \geq & f((1-\theta)\vec{u} + \theta\vec{v}) \\ \equiv & (1-\theta)f(\vec{u}) + \theta f(\vec{v})) - f((1-\theta)\vec{u} + \theta\vec{v}) & \geq & 0 \end{array}$$

We can reason that

$$
\begin{aligned}
& (1-\theta)f(\vec{u}) + \theta f(\vec{v})) - f((1-\theta)\vec{u} + \theta\vec{v}) \\
= \;& (1-\theta)(\vec{u}^T K \vec{u}) + \theta(\vec{v}^T K \vec{v}) - [(1-\theta)\vec{u} + \theta\vec{v}]^T K [(1-\theta)\vec{u} + \theta\vec{v}] \\
= \;& (1-\theta)(\vec{u}^T K \vec{u}) + \vec{v}^T K \vec{v} - (1-\theta)(\vec{v}^T K \vec{v}) - [(1-\theta)\vec{u} + \theta\vec{v}]^T K [(1-\theta)\vec{u} + \theta\vec{v}] \\
= \;& (1-\theta)(\vec{u}^T K \vec{u}) + \vec{v}^T K \vec{v} - (1-\theta)(\vec{v}^T K \vec{v})w \\
& -(1-\theta)^2(\vec{u}^T K \vec{u}) - \theta^2(\vec{v}^T K \vec{v}) - 2(1-\theta)\theta(\vec{u}^T K \vec{v}) \\
= \;& (1-\theta)\theta(\vec{u}^T K \vec{u}) + \vec{v}^T K \vec{v} - (1-\theta)(\vec{v}^T K \vec{v}) \\
& -\theta^2(\vec{v}^T K \vec{v}) - 2(1-\theta)\theta(\vec{u}^T K \vec{v}) \\
= \;& (1-\theta)\theta(\vec{u}^T K \vec{u}) + \vec{v}^T K \vec{v} - (1-\theta)(\vec{v}^T K \vec{v}) \\
& -(1 - 2(1-\theta) + (1-\theta)^2)(\vec{v}^T K \vec{v}) - 2(1-\theta)\theta(\vec{u}^T K \vec{v}) \\
= \;& (1-\theta)\theta(\vec{u}^T K \vec{u}) + \theta\vec{v}^T K \vec{v} \\
& -(1 - 2(1-\theta) + (1-\theta)^2)(\vec{v}^T K \vec{v}) + 2(1-\theta)\theta(\vec{u}^T K \vec{v}) \\
= \;& (1-\theta)\theta(\vec{u}^T K \vec{u}) + ((1-\theta) - (1-\theta)^2)\vec{v}^T K \vec{v} \\
& -2(1-\theta)\theta(\vec{u}^T K \vec{v}) \\
= \;& (1-\theta)\theta(\vec{u}^T K \vec{u}) + (1-\theta)\theta\vec{v}^T K \vec{v} - 2(1-\theta)\theta(\vec{u}^T K \vec{v}) \\
= \;& (1-\theta)\theta(\vec{u}^T K \vec{u} - 2(\vec{u}^T K \vec{v}) + \vec{v}^T K \vec{v}) \\
= \;& (1-\theta)\theta[\vec{u} - \vec{v}]^T K [\vec{u} - \vec{v}] \\
\geq \;& (1-\theta)\theta\, 0 \\
\geq \;& 0
\end{aligned}
$$

Observation: The above reasoning holds for $K \succeq 0$, so the quadratic form of a symmetric positive semidefinite matrix is also a convex function.

Theorem: For any convex $f : \mathbb{R}^n \to \mathbb{R}$, and any nonnegative $l \in \mathbb{R}_+$, the composition

$$(lf) : \mathbb{R}^n \to \mathbb{R} \tag{18.20}$$

is a convex function.

Proof: Assume that $\theta \in \mathbb{R}_+$, that $\theta \leq 1$, and that $l \in \mathbb{R}_+$. Using Definition 18.2, and Property 18.7,

$$
\begin{aligned}
& (f((1-\theta)\vec{u} + \theta\vec{v})) && \leq ((1-\theta)f(\vec{u}) + \theta f(\vec{v}))) \\
\to \;& (lf((1-\theta)\vec{u} + \theta\vec{v})) && \leq ((1-\theta)lf(\vec{u}) + \theta lf(\vec{v})))
\end{aligned}
$$

so $(lf) : \mathbb{R}^n \to \mathbb{R}$ is a convex function.

Theorem: For any $f_A : \mathbb{R}^n \to \mathbb{R}$ that is a convex function, and any $f_B : \mathbb{R}^n \to \mathbb{R}$ that is a convex function, the composition

$$(f_A + f_B) : \mathbb{R}^n \to \mathbb{R} \tag{18.21}$$

is a convex function.

Proof: Assume that $\theta \in \mathbb{R}_+$ and that $\theta \leq 1$. Using Definition 18.2, and Property 18.7. Then

$$\begin{array}{rrcl} & ((f_A((1-\theta)\vec{u} + \theta\vec{v})) & \leq & ((1-\theta)f_A(\vec{u}) + \theta f_A(\vec{v}))) \\ \text{and} & (f_B((1-\theta)\vec{u} + \theta\vec{v})) & \leq & ((1-\theta)f_B(\vec{u}) + \theta f_B(\vec{v})))) \\ \to & ((f_A + f_B)((1-\theta)\vec{u} + \theta\vec{v})) & \leq & ((1-\theta)(f_A + f_B)(\vec{u}) + \theta(f_A + f_B)(\vec{v}))) \end{array}$$

so $(f_A + f_B) : \mathbb{R}^n \to \mathbb{R}$ is a convex function.

Theorem: For any $f : \mathbb{R}^m \to \mathbb{R}$ that is a convex function, any matrix $M \in \mathbb{R}^m \times \mathbb{R}^n$, any vector $\vec{c} \in \mathbb{R}^m$, and any vector $\vec{w} \in \mathbb{R}^n$, the composition

$$(g : \mathbb{R}^n \to \mathbb{R}) \stackrel{\text{def}}{=} f(M\vec{w} + \vec{c}) \tag{18.22}$$

is a convex function.

Proof: Assume that $\theta \in \mathbb{R}_+$ and that $\theta \leq 1$. From Definition 18.2,

$$f((1-\theta)\vec{u} + \theta\vec{v}) \leq ((1-\theta)f(\vec{u}) + \theta f(\vec{v}))$$

Substituting $M\vec{w} + \vec{c}$ for $\vec{u}$ and $\vec{v}$ in the definition is

$$\begin{array}{rcl} f((1-\theta)(M\vec{u} + \vec{c}) + \theta(M\vec{v} + \vec{c})) & \leq & ((1-\theta)f(M\vec{u} + \vec{c}) + \theta f(M\vec{v} + \vec{c})) \\ \equiv \quad g((1-\theta)\vec{u} + \theta\vec{v}) & \leq & ((1-\theta)f(\vec{u}) + \theta f(\vec{v})) \end{array}$$

so $g(\vec{w})$ is a convex function.

18.5.3 Theorems for Convex Sets

Theorem: the intersection of two convex sets is a convex set.

For any convex set $\mathbb{V}_A \subseteq \mathbb{R}^n$ and any convex set $\mathbb{V}_B \subseteq \mathbb{R}^n$, the intersection

$$\mathbb{V}_A \cap \mathbb{V}_B \tag{18.23}$$

is a convex set.

Proof: Consider any $\vec{u}:((\vec{u} \in \mathbb{V}_A) \wedge (\vec{u} \in \mathbb{V}_B))$, any $\vec{v}:((\vec{v} \in \mathbb{V}_A) \wedge (\vec{v} \in \mathbb{V}_B))$, and any $\theta \in \mathbb{R}_+$ such that $\theta \leq 1$. Then

$$\begin{aligned}
\mathbb{V}_A \text{ is convex} &\rightarrow ((1-\theta)\vec{u} + \theta\vec{v}) \in \mathbb{V}_A \\
\mathbb{V}_B \text{ is convex} &\rightarrow ((1-\theta)\vec{u} + \theta\vec{v}) \in \mathbb{V}_B \\
(\mathbb{V}_A \text{ is convex}) \wedge (\mathbb{V}_B \text{ is convex}) &\rightarrow (((1-\theta)\vec{u} + \theta\vec{v}) \in \mathbb{V}_A) \wedge (((1-\theta)\vec{u} + \theta\vec{v}) \in \mathbb{V}_B) \\
&\rightarrow (((1-\theta)\vec{u} + \theta\vec{v}) \in (\mathbb{V}_A \cap \mathbb{V}_B)
\end{aligned}$$

so $\mathbb{V}_A \cap \mathbb{V}_B$ is convex.

Theorem: an affine transformation of a convex set is convex.

> For any convex set $\mathbb{V} \subseteq \mathbb{R}^n$, any vector $\vec{w} \in \mathbb{V}$, any matrix $M \in \mathbb{R}^m \times \mathbb{R}^n$, and any vector $\vec{c} \in \mathbb{R}^m$, the set
>
> $$\mathbb{V}_A \stackrel{\text{def}}{=} \{\vec{u}:((\vec{u} \in \mathbb{R}^m) \wedge (\vec{u} = M\vec{w} + \vec{c}))\} \quad \text{is convex} \tag{18.24}$$

Proof: For any vector $\vec{a} \in \mathbb{V}$, construct $\vec{a}_A = M\vec{a}+\vec{c}$. For any vector $\vec{b} \in \mathbb{V}$, construct $\vec{b}_A = M\vec{b}+\vec{c}$. Because $\mathbb{V}$ is a convex set, the vector $(1-\theta)\vec{a} + \theta\vec{b}$ is in the set $\mathbb{V}$.

By construction, $\vec{a}_A \in \mathbb{R}^m$ and $\vec{b}_A \in \mathbb{R}^m$ so $\vec{u} = (1-\theta)\vec{a}_A + \theta\vec{b}_A$ is in $\mathbb{R}^m$. Then

$$\begin{aligned}
M[(1-\theta)\vec{a} + \theta\vec{b}] + \vec{c} &= (1-\theta)[M\vec{a} + \vec{c}] + \theta[M\vec{b} + \vec{c}] \\
&= (1-\theta)M\vec{a} + \theta M\vec{b} + ((1-\theta) + \theta)\vec{c} \\
&= M[(1-\theta)\vec{a} + \theta\vec{b}] + \vec{c} \\
&= (1-\theta)\vec{a}_A + \theta\vec{b}_A \\
&= \vec{u}
\end{aligned}$$

so $\vec{u} \in \mathbb{V}_A$.

End of Extra Notes

References

[1] Antoniou A, Lu WS: Practical Optimization: Algorithms and Engineering Applications. Springer Science & Business Media, 2007

[2] Beck A: Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB. Siam Press, 2014

[3] Boyd S, Vandenberghe L: Convex Optimization. Cambridge University Press, 2004