Optimization

Timur Musin

University of Freiburg

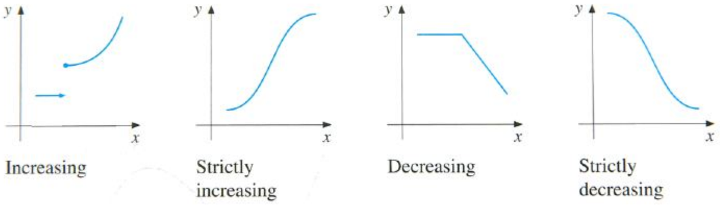
October 14, 2014

Increasing and Decreasing Functions

Consider a function $f(x)$ that is defined on an interval I and let $x_1, x_2 \in I$, then:

- If $f(x_2) \geq f(x_1)$ whenever $x2 > x1$, then $f(x)$ is **increasing** in I.
- If $f(x_2)\ f(x_1)$ whenever $x2 > x1$, then $f(x)$ is **strictly increasing** in I.
- If $f(x_2) \leq f(x_1)$, then $f(x)$ is **decreasing** in I.
- If $f(x_2) < f(x_1)$, then $f(x)$ is **strictly decreasing** in I.

Graphically:

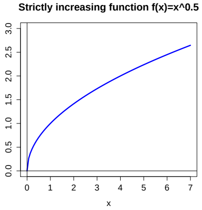


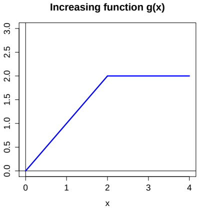
Increasing and Decreasing Functions

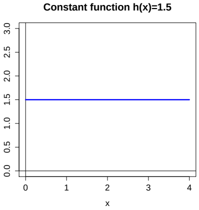
Consider a differentiable function $f(x)$ and an interval I from its domain. We can obtain the following information regarding the function from its first derivative:

- If $f'(x) \geq 0$ for all $x \in I$ then $f(x)$ is **increasing** in I.
- If $f'(x) \geq 0$ for all $x \in I$ then $f(x)$ is **strictly increasing** in I.
- If $f'(x) \leq 0$ for all $x \in I$ then $f(x)$ is **decreasing** in I.
- If $f'(x) \leq 0$ for all $x \in I$ then $f(x)$ is **strictly decreasing** in I.
- If $f'(x) = 0$ for all $x \in I$ then $f(x)$ is **constant** in I.

Graphically:





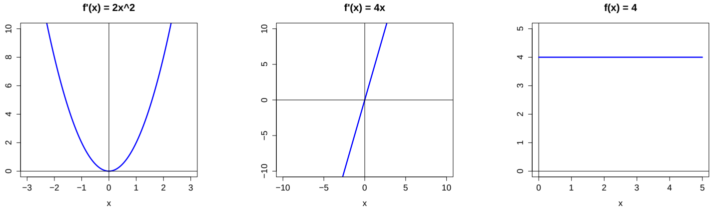


Convex and Concave Functions

Consider a twice-differentiable function $f(x)$ and an interval I from its domain. We can obtain the following information regarding the shape of the function from its second derivative:

- If $f''(x) \geq (>)\ 0$ for all $x \in I$ then $f'(x)$ is an increasing function in I and $f(x)$ is a **convex (strictly convex)** function in I.

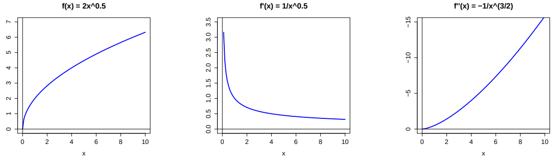
Graphically:



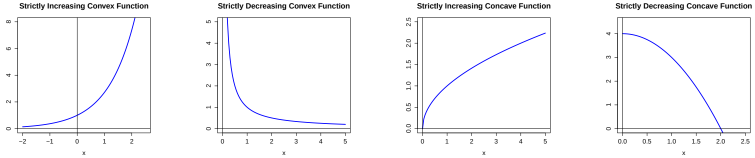
Convex and Concave Functions

- If $f''(x) \leq (<)\ 0$ for all $x \in I$ then $f'(x)$ is a decreasing function in I and $f(x)$ is a **concave (strictly concave)** function in I.

Graphically:



- Examples:



Convex Linear Combinations of Vectors

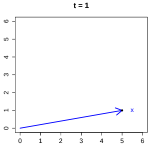
Consider the vectors $x = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ and $y = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$. For $t \in [0, 1]$ draw the linear combinations of the vectors $\alpha * x + (1 - \alpha) * y$.

Let's represent graphically the position of the linear combination of the given vectors, $t * x + (1 - t) * y$, for $t = 1$, $t = 0$ and $t = 0.5$.

- Let $t = 1$; then

  $$t * x + (1 - t) * y = 1 * x + (1 - 1) * y = x = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

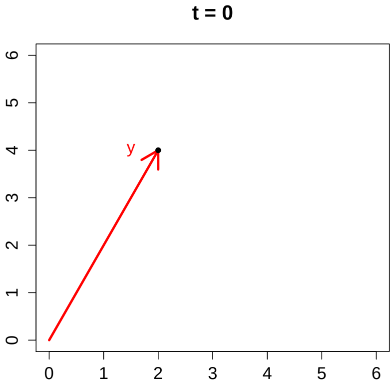
Graphically:



Convex Linear Combinations of Vectors

- Let $t = 0$; then
  $$t * x + (1 - t) * y = 0 * x + (1 - 0) * y = y = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

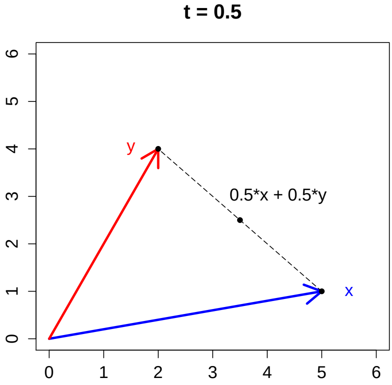
Graphically:



Convex Linear Combinations of Vectors

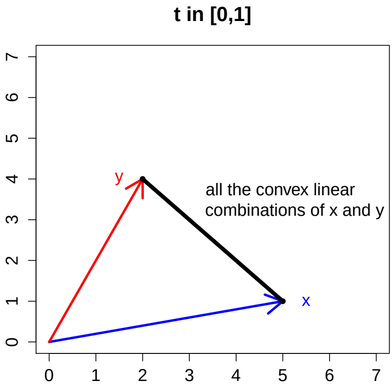
- Let $t = 0.5$; then $t * x + (1 - t) * y =$
$0.5 * x + (1 - 0.5) * y = 0.5 * \begin{bmatrix} 5 \\ 1 \end{bmatrix} + 0.5 * \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3.5 \\ 2.5 \end{bmatrix}$.

Graphically:



Convex Linear Combinations of Vectors

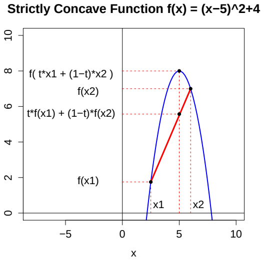
All the possible linear combinations of the given vectors x and y for $t \in [0, 1]$ will fill in the entire dashed line (convince yourself that this is really the case: try $t = -2,\ -1,\ 0.2,\ 0.6,\ 1,\ 2$). Graphically all the linear combinations of any two vectors for $t \in [0, 1]$ represent a line segment connecting the endpoints of these vectors:



More on Convex and Concave Functions

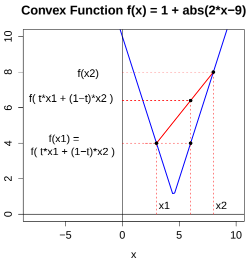
More general definitions of concave and convex functions:

- A function $f(x)$ is called **concave** if any line segment connecting any two dots on the graph is either **below** or on the graph.
- **Mathematically:**
  if $f: \mathbb{R} \to \mathbb{R}$ and $f(tx_1 + (1-t)x_2) \geq tf(x_1) + (1-t)f(x_2)$ holds for any $t \in [0,1]$ in an interval I then $f(x)$ is **concave** in I.
- **Graphically:**



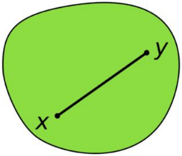
More on Convex and Concave Functions

- A function $f(x)$ is called **convex** if any line segment connecting any two dots on the graph is either **above** or on the graph.
- **Mathematically:**
  if $f : \mathbb{R} \rightarrow \mathbb{R}$ and $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$ holds for any $t \in [0,1]$ in an interval I then $f(x)$ is **convex** in I.
- **Graphically:**

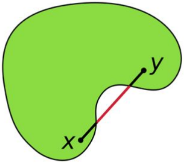


Quasi-Concave and Quasi-Convex Functions

- An object is said to be **convex** if for every pair of points within the object, every point on the straight line segment that connects the points is also within the object.
- Illustration of a **convex** set:

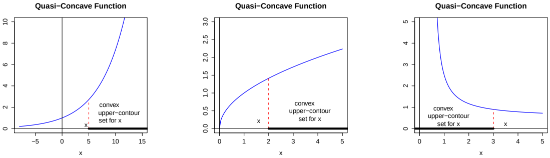


- Illustration of a **non-convex** set:



Quasi-Concave and Quasi-Convex Functions

- A function $f : \mathbb{R} \to \mathbb{R}$ is **quasi-concave** if its upper-contour set $\{y \in \mathbb{R} : f(y) \geq f(x)\}$ is convex.
- **Equivalently:** a function $f : \mathbb{R} \to \mathbb{R}$ is **quasi-concave** if $f(tx + (1-t)y) \geq min\{f(x), f(y)\}$.
- Examples:

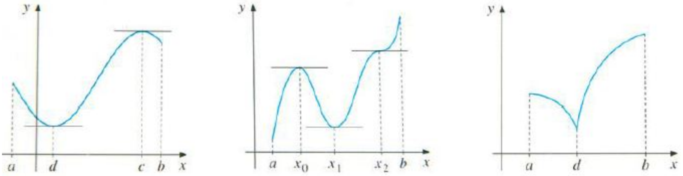


Critical Points. Necessary First-Order Conditions.

- **Definition:** a **critical** point of a function $f(x)$ is the number c in its domain where either $f'(c) = 0$ or $f'(c)$ doesn't exist.

- Suppose that a function $f(x)$ is differentiable in an interval I and that c is an interior point in I. A **necessary condition** for $x = c$ to be a maximum or minimum (extremum) point of $f(x)$ is $f'(c) = 0$.

Graphically:



First Derivative Test For Local Extrema Points

Suppose c is a stationary point of $f(x)$.

- If $f'(x) \geq 0$ throughout some interval (a, c) to the left of c and $f'(x) \leq 0$ throughout some interval (c, b) to the right of c, then $x = c$ is a local maximum point of f.
- If $f'(x) \leq 0$ throughout some interval (a, c) to the left of c and $f'(x) \geq 0$ throughout some interval (c, b) to the right of c, then $x = c$ is a local minimum point of f.
- If $f'(x) < 0$ (or $f'(x) > 0$) both throughout some interval (a, b) to the left of c and hroughout some interval (c, b) to the right of c, then $x = c$ is not a local extreme point.

Second-Derivative Test For Local Extrema Points

Let $f(x)$ be a twice differentiable function in an interval I and $c \in I$, then:

- $f'(c) = 0$ and $f''(c) < 0 \Rightarrow c$ is a strict local maximum point.
- $f'(c) = 0$ and $f''(c) > 0 \Rightarrow c$ is a strict local minimum point.
- $f'(c) = 0$ and $f''(c) = 0 \Rightarrow$ nothing can be said about the nature of the point (see examples below).

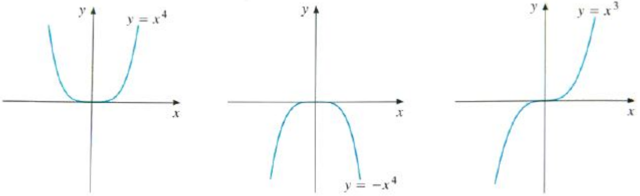


Figure 2 $f'(0) = f''(0) = 0$, and 0 is a minimum point

Figure 3 $f'(0) = f''(0) = 0$, and 0 is a maximum point

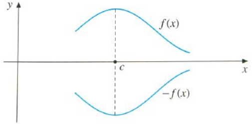
Figure 4 $f'(0) = f''(0) = 0$, and 0 is an inflection point

Global Maximum/Minimum of a Function

Consider a function $f(x)$ with a domain $\mathbb{D}$:

- **Definition:** $c \in \mathbb{D}$ is a **global maximum** point of $f(x) \Leftrightarrow f(x) \leq f(c)$ for all $x \in \mathbb{D}$.
- **Definition:** $d \in \mathbb{D}$ is a **global minimum** point of $f(x) \Leftrightarrow f(x) \geq f(d)$ for all $x \in \mathbb{D}$.
- If $f(x) < f(c)$ or $f(x) > f(d)$ for all $x \in \mathbb{D}$ we say that c and d are **strict global maximum/minimum** respectively.

Graphically:



The point c is a global maximum point for $f(x)$ and a global minimum point for $-f(x)$.

Global Maximum/Minimum of a Function

To find a global (absolute) maximum/minimum value of $f(x)$ on a closed interval $[a, b]$:

1. Evaluate $f(x)$ at the critical points in (a, b).
2. Evaluate $f(x)$ at a and b.
3. Global maximum is the maximum of steps 1) and 2).
   Global minimum is the minimum of steps 1) and 2).

Constrained Optimization: Equality Constraints

Let $f(x, y)$ and $g(x, y)$ be twice differentiable functions. Consider the following constrained optimization problem:

$$\begin{cases} \max\limits_{x,y} \; f(x, y) \\ s.t. \; g(x, y) = c \end{cases}$$

In order to solve this optimization problem with an equality constraint, we will use the **Augmented Lagrangian method**:

1. Construct the Lagrangian:

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda \cdot [c - g(x, y)]$$

Constrained Optimization: Equality Constraints

2. Take the partial derivatives of $\mathcal{L}(x, y, \lambda)$ with respect to all the independent variables:

$$\begin{cases} \dfrac{\partial \mathcal{L}(x^*, y^*, \lambda^*)}{\partial x} = \dfrac{\partial f}{\partial x}(x^*, y^*) - \lambda^* \cdot \dfrac{\partial g}{\partial x}(x^*, y^*) & = 0 \quad (1) \\ \dfrac{\partial \mathcal{L}(x^*, y^*, \lambda^*)}{\partial y} = \dfrac{\partial f}{\partial y}(x^*, y^*) - \lambda^* \cdot \dfrac{\partial g}{\partial y}(x^*, y^*) & = 0 \quad (2) \\ \dfrac{\partial \mathcal{L}(x^*, y^*, \lambda^*)}{\partial \lambda} = c - g(x^*, y^*) & = 0 \quad (3) \end{cases}$$

Constrained Optimization: Equality Constraints

3. Solve the system of the obtained equations following the indicated steps:
   - from (1) and (2):

$$\lambda^* = \frac{\frac{\partial f}{\partial x}(x^*, y^*)}{\frac{\partial g}{\partial x}(x^*, y^*)} = \frac{\frac{\partial f}{\partial y}(x^*, y^*)}{\frac{\partial g}{\partial y}(x^*, y^*)} \tag{4}$$

   - from (3) and (4):

$$\textit{express } x^*, y^* \tag{5}$$

   - from (5) and (1) or (2):

$$\textit{express } \lambda^* \tag{6}$$

Constrained Optimization: Equality Constraints

Interpreting λ:

- so far having started with a constrained optimization problem

  $$\begin{cases} \max_{x,y} \; f(x,y) \\ s.t. \; g(x,y) = c \end{cases}$$

  and having constructed the Lagrangian

  $\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda \cdot [c - g(x,y)]$, we obtained the optimal values of x^*, y^*, λ^*.
- Now suppose that c is a parameter and that $x^* = x^*(c)$, $y^* = y^*(c)$, $\lambda^*(c)$; we can then introduce $F(c) := f(x^*(c), y^*(c))$.
- $F(c)$ is interpreted as a value function that determines the maximum level of $f(x^*(c), y^*(c))$ for any given c.

Constrained Optimization: Equality Constraints

- We are now interested in $\frac{dF(c)}{dc} = ?$
- Taking the total derivative of $F(x)$, we obtain

$$\frac{dF(c)}{dc} = \frac{\partial f(x^*(c), y^*(c))}{\partial x} \cdot \frac{dx^*(c)}{dc} + \frac{\partial f(x^*(c), y^*(c))}{\partial y} \cdot \frac{dy^*(c)}{dc} =$$

$$\underbrace{\lambda^*(c) \cdot \frac{\partial g(x^*(c), y^*(c))}{\partial x}}_{\text{from (1)}} \cdot \frac{dx^*(c)}{dc} +$$

$$\underbrace{\lambda^*(c) \cdot \frac{\partial g(x^*(c)), y^*(c))}{\partial y}}_{\text{from (2)}} \cdot \frac{dy^*(c)}{dc} =$$

$$= \lambda^*(c) \cdot \underbrace{\left[\frac{\partial g(x^*(c), y^*(c))}{\partial x} \cdot \frac{dx^*(c)}{dc} + \frac{\partial g(x^*(c), y^*(c))}{\partial y} \cdot \frac{dy^*(c)}{dc} \right]}_{?}$$

Constrained Optimization: Equality Constraints

Simplifying the expression above the "?":

- Substituting $x^*(c)$, $y^*(c)$ into the original constraint $g(x, y) = c$, we obtain $g(x^*(c), y^*(c)) = c$.
- Differentiating both parts of the equation with respect to c, we get:

$$\frac{\partial g(\cdot)}{\partial x} \cdot \frac{dx^*(c)}{dc} + \frac{\partial g(\cdot)}{\partial y} \cdot \frac{dy^*(c)}{dc} = \frac{dc}{dc} = 1$$

- Therefore, we can conclude that

$$\frac{dF(c)}{dc} \equiv \frac{df(x^*(c), y^*(c))}{dc} = \lambda^*(c)$$

.

- $\lambda^*(c)$, thus, shows by how much the optimal value of the objective function changes if we change we value of the constraint.

Constrained Optimization: Inequality Constraints

- We will be facing one of two types of constraints:
  1. **Binding:** $g(x^*(c), y^*(c)) = c$
  2. **Slack:** $g(x^*(c), y^*(c) < c$

- Therefore, consider a different optimization problem:

$$\begin{cases} \max\limits_{x,y} \; f(x,y) \\ s.t. \; g(x,y) \leq c \end{cases}$$

Constrained Optimization: Inequality Constraints

- The solution procedure is similar to the one introduced above. However, the system of equations is supplemented with the so-called **Kuhn-Tucker** conditions (9) – (11):

$$
\begin{cases}
\dfrac{\partial \mathcal{L}(x^*, y^*, \lambda^*)}{\partial x} = 0 & (7) \\
\dfrac{\partial \mathcal{L}(x^*, y^*, \lambda^*)}{\partial y} = 0 & (8) \\
c - g(x^*, y^*) \geq 0 & (9) \\
\lambda^* \geq 0 & (10) \\
\lambda^* \cdot [c - g(x^*, y^*)] = 0 & (11)
\end{cases}
$$

Constrained Optimization: Inequality Constraints

There are two possible cases:

1. If $\lambda^* = 0$ then $c - g(x^*, y^*) > 0$, i.e. the constraint is **slack** and we can continue with **Unconstrained Optimization**.
2. If $\lambda^* > 0$ then $c - g(x^*, y^*) = 0$, i.e. the constraint is **binding** and we should use the **Lagrangian**.

Kuhn-Tucker conditions act as a "switcher" between Unconstrained and Constrained Optimization.