

Course notes for MATH 524: Non-Linear Optimization

Francisco Blanco-Silva

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTH CAROLINA

E-mail address: `blanco@math.sc.edu`

URL: `people.math.sc.edu/blanco`

Contents

List of Figures	v
Chapter 1. Review of Optimization from Vector Calculus	1
1. The Theory of Optimization	4
Exercises	5
Chapter 2. Existence of Extrema	7
1. Continuous functions on compact domains	7
2. Continuous functions on unbounded domains	7
3. Convex functions	8
Exercises	9
Chapter 3. Characterization of Extrema	11
Chapter 4. Nonlinear optimization	13
Bibliography	15

List of Figures

1.1	Details of the graph of $\mathcal{R}_{1,1}$	2
1.2	Global minima in unbounded domains	4
1.3	Contour plots for problem 1.4	6
2.1	Convex Functions.	9

CHAPTER 1

Review of Optimization from Vector Calculus

The starting point of these notes is the concept of *optimization* as developed in MATH 241 (see e.g. [1, Chapter 14])

DEFINITION. Let $D \subseteq \mathbb{R}^2$ be a region on the plane containing the point (x_0, y_0) . We say that the real-valued function $f: D \rightarrow \mathbb{R}$ has a *local minimum* at (x_0, y_0) if $f(x_0, y_0) \leq f(x, y)$ for all domain points (x, y) in an open disk centered at (x_0, y_0) . In that case, we also say that $f(x_0, y_0)$ is a *local minimum value* of f in D .

Emphasis was made to find conditions on the function f to guarantee existence and identification of minima:

THEOREM 1.1. Let $D \subseteq \mathbb{R}^2$ and let $f: D \rightarrow \mathbb{R}$ be a function for which first partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist in D . If $(x_0, y_0) \in D$ is a local minimum of f , then $\nabla f(x_0, y_0) = 0$.

The local minima of these functions are among the zeros of the equation $\nabla f(x, y) = 0$, the so-called *critical points* of f . More formally:

DEFINITION. An interior point of the domain of a function $f(x, y)$ where both directional derivatives are zero, or where at least one of the directional derivatives do not exist, is a *critical point* of f .

In order to select *some* minima, we employed the *Second Derivative Test for Local Extreme Values*

THEOREM 1.2. Suppose that $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ and its first and second partial derivatives are continuous throughout a disk centered at the point (x_0, y_0) , and that $\nabla f(x_0, y_0) = 0$. Then $f(x_0, y_0)$ is a local minimum value if the two following conditions are satisfied:

$$\begin{aligned}
 (1) \quad & \frac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0 \\
 (2) \quad & \det \underbrace{\begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(x_0, y_0) & \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) \\ \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) & \frac{\partial^2 f}{\partial y^2}(x_0, y_0) \end{bmatrix}}_{\text{Hess}f(x_0, y_0)} > 0
 \end{aligned}$$

REMARK 1.1. The restriction of this result to univariate functions is even simpler: Suppose f'' is continuous on an open interval that contains x_0 . If $f'(x_0) = 0$ and $f''(x_0) > 0$, then f has a local minimum at x_0 .

EXAMPLE 1.1 (Rosenbrock Functions). Given strictly positive parameters $a, b > 0$, consider the (a, b) -Rosenbrock function

$$\mathcal{R}_{a,b}(x, y) = (a - x)^2 + b(y - x^2)^2.$$

It is easy to see that Rosenbrock functions are polynomials (prove it!). The domain is therefore the whole plane. Figure 1.1 illustrates a contour plot with several level lines of $\mathcal{R}_{1,1}$ on the domain $D = [-2, 2] \times [-1, 3]$, as well as its graph.

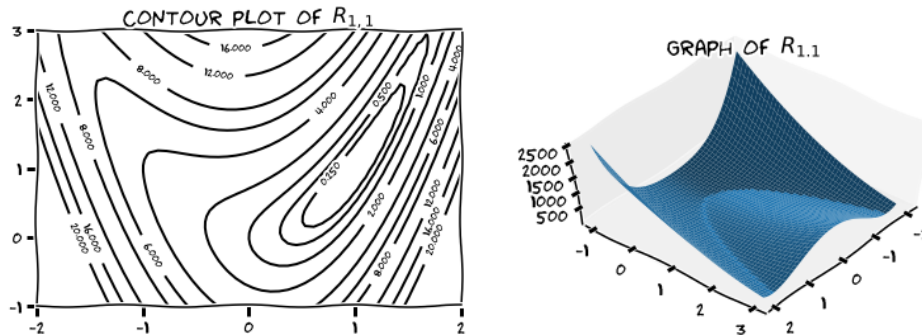


FIGURE 1.1. Details of the graph of $\mathcal{R}_{1,1}$

It is also easy to verify that the image is the interval $[0, \infty)$. Indeed, note first that $\mathcal{R}_{a,b}(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$. Zero is attained: $\mathcal{R}_{a,b}(a, a^2) = 0$. Note also that $\mathcal{R}_{a,b}(0, y) = a^2 + by^2$ is a polynomial of degree 2, therefore unbounded.

Let's locate all local minima:

- The gradient and Hessian are given respectively by

$$\nabla \mathcal{R}_{a,b}(x, y) = [2(x - a) + 4bx(x^2 - y), b(y - x^2)]$$

$$\text{Hess} \mathcal{R}_{a,b}(x, y) = \begin{bmatrix} 12bx^2 - 4by + 2 & -4bx \\ -4bx & 2b \end{bmatrix}$$

- The search for critical points $\nabla \mathcal{R}_{a,b} = \mathbf{0}$ gives only the point (a, a^2) .
- $\frac{\partial^2 \mathcal{R}_{a,b}}{\partial x^2}(a, a^2) = 8ba^2 + 2 > 0$.
- The Hessian at that point has positive determinant:

$$\det \text{Hess} \mathcal{R}_{a,b}(a, a^2) = \det \begin{bmatrix} 8ba^2 + 2 & -4ab \\ -4ab & 2b \end{bmatrix} = 4b > 0$$

There is only one local minimum at (a, a^2) .

The second step was the notion of *global (or absolute) minima*: points (x_0, y_0) that satisfy $f(x_0, y_0) \leq f(x, y)$ for any point (x, y) in the domain of f . We always started with the easier setting, in which we placed restrictions on the domain of our functions:

THEOREM 1.3. *A continuous real-valued function always attains its minimum value on a compact set K . To search for global minima, we perform the following steps:*

Interior Candidates: *List the critical points of f located in the interior of K .*

Boundary Candidates: *List the points in the boundary of K where f may have minimum values.*

Evaluation/Selection: *Evaluate f at all candidates and select the one(s) with the smallest value.*

EXAMPLE 1.2. A flat circular plate has the shape of the region $x^2 + y^2 \leq 1$. The plate, including the boundary, is heated so that the temperature at the point (x, y) is given by $f(x, y) = 100(x^2 + 2y^2 - x)$ in Celsius degrees. Find the temperature at the coldest point of the plate.

We start by searching for critical points. The equation $\nabla f(x, y) = 0$ gives $x = \frac{1}{2}$, $y = 0$. The point $(\frac{1}{2}, 0)$ is clearly inside of the plate. This is our first candidate.

The border of the plate can be parameterized by $\varphi(t) = (\cos t, \sin t)$ for $t \in [0, 2\pi)$. The search for minima in the boundary of the plate can then be coded as an optimization problem for the function $h(t) = (f \circ \varphi)(t) = 100(\cos^2 t + 2\sin^2 t - \cos t)$ on the interval $[0, 2\pi)$. Note that $h'(t) = 0$ for $t \in \{0, \frac{2}{3}\pi\}$ in $[0, 2\pi)$. We thus have two more candidates:

$$\varphi(0) = (1, 0) \quad \varphi(\tfrac{2}{3}\pi) = (-\tfrac{1}{2}, \tfrac{1}{2}\sqrt{3})$$

Evaluation of the function at all candidates gives us the solution to this problem:

$$f(\tfrac{1}{2}, 0) = -25^\circ\text{C}.$$

On a second setting, we remove the restriction of boundedness of the function. In this case, global minima will only be guaranteed for very special functions.

EXAMPLE 1.3. Any polynomial $p_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ with even degree $n \geq 2$ and positive leading coefficient satisfies $\lim_{|x| \rightarrow \infty} p_n(x) = +\infty$. To see this, we may write

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 = a_n x^n \left(1 + \frac{a_{n-1}}{a_n x} + \cdots + \frac{a_0}{a_n x^n}\right)$$

The behavior of each of the factors as the absolute value of x goes to infinity leads to our claim.

$$\begin{aligned} \lim_{|x| \rightarrow \infty} a_n x^n &= +\infty, \\ \lim_{|x| \rightarrow \infty} \left(1 + \frac{a_{n-1}}{a_n x} + \cdots + \frac{a_0}{a_n x^n}\right) &= 1. \end{aligned}$$

It is clear that a polynomial of this kind must attain a minimum somewhere in its domain. The critical points will lead to them.

EXAMPLE 1.4. Find the global minima of the function $f(x) = \log(x^4 - 2x^2 + 2)$ in \mathbb{R} .

Note first that the domain of f is the whole real line, since $x^4 - 2x^2 + 2 = (x^2 - 1)^2 + 1 \geq 1$ for all $x \in \mathbb{R}$. Note also that we can write $f(x) = (g \circ h)(x)$ with $g(x) = \log(x)$ and $h(x) = x^4 - 2x^2 + 1$. Since g is one-to-one and increasing, we can focus on h to obtain the requested solution. For instance, $\lim_{|x| \rightarrow \infty} f(x) = +\infty$, since $\lim_{|x| \rightarrow \infty} h(x) = +\infty$. This guarantees the existence of global minima. To look for it, h again points to the possible locations by solving for its critical points: $h'(x) = 0$. We have then that f attains its minima at $x = \pm 1$.

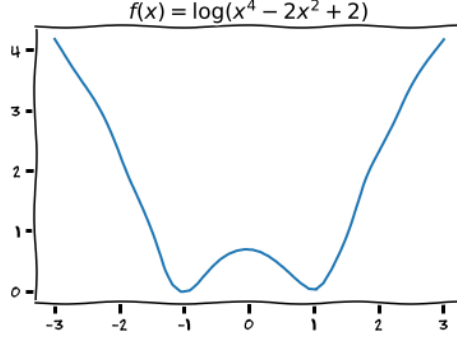


FIGURE 1.2. Global minima in unbounded domains

1. The Theory of Optimization

The purpose of these notes is the development of a theory to deal with optimization in a more general setting.

- We start in an Euclidean d -dimensional space with the usual topology based on the distance

$$\|\mathbf{x} - \mathbf{y}\| = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle^{1/2} = \sqrt{\sum_{k=1}^d (x_k - y_k)^2}.$$

- Given a real-valued function $f: D \rightarrow \mathbb{R}$ on a domain $D \subseteq \mathbb{R}^d$, we define the concept of *extrema*:

DEFINITION. Given a set $D \subseteq \mathbb{R}^d$, and a real-valued function $f: D \rightarrow \mathbb{R}$, we say that a point $\mathbf{x}^* \in D$ is:

- A *global minimum* for f on D if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$.
- A *global maximum* for f on D if $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in D$.
- A *strict global minimum* for f on D if $f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\mathbf{x} \in D \setminus \{\mathbf{x}^*\}$.
- A *strict global maximum* for f on D if $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \in D \setminus \{\mathbf{x}^*\}$.
- A *local minimum* for f on D if there exists $\delta > 0$ so that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in B_\delta(\mathbf{x}^*) \cap D$.
- A *local maximum* for f on D if there exists $\delta > 0$ so that $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in B_\delta(\mathbf{x}^*) \cap D$.
- A *local minimum* for f on D if there exists $\delta > 0$ so that $f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\mathbf{x} \in B_\delta(\mathbf{x}^*) \cap D$, $\mathbf{x} \neq \mathbf{x}^*$.
- A *local maximum* for f on D if there exists $\delta > 0$ so that $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \in B_\delta(\mathbf{x}^*) \cap D$, $\mathbf{x} \neq \mathbf{x}^*$.

In this setting, the objective of *optimization* is the following program:

Existence of extrema: Establish results that guarantee the existence of extrema depending on the properties of D and f . We include the development of these ideas in chapter 2.

Characterization of extrema: Establish results that describe conditions for points $\mathbf{x} \in D$ to be extrema of f . We develop these in chapter 3.

Tracking extrema: Design robust numerical algorithms that find the extrema for scientific computing purposes. This is the core of these notes. The presentation of this topic is done in chapter 4.

Exercises

PROBLEM 1.1. Develop similar statements as in Definition 1, Theorems 1.1, 1.2 and 1.3, but for *local* and *global maxima*.

PROBLEM 1.2 (Domains). Find and sketch the domain of the following functions.

- (a) $f(x, y) = \sqrt{y - x - 2}$
- (b) $f(x, y) = \log(x^2 + y^2 - 4)$
- (c) $f(x, y) = \frac{(x-1)(y+2)}{(y-x)(y-x^3)}$
- (d) $f(x, y) = \log(xy + x - y - 1)$

PROBLEM 1.3 (Contour plots). Find and sketch the level lines $f(x, y) = c$ on the same set of coordinate axes for the given values of c .

- (a) $f(x, y) = x + y - 1$, $c \in \{-3, -2, -1, 0, 1, 2, 3\}$.
- (b) $f(x, y) = x^2 + y^2$, $c \in \{0, 1, 4, 9, 16, 25\}$.
- (c) $f(x, y) = xy$, $c \in \{-9, -4, -1, 0, 1, 4, 9\}$

PROBLEM 1.4. Use a Computer Algebra System of your choice to produce contour plots of the given functions on the given domains.

- (a) $f(x, y) = (\cos x)(\cos y)e^{-\sqrt{x^2+y^2}/4}$ on $[-2\pi, 2\pi] \times [-2\pi, 2\pi]$.
- (b) $g(x, y) = \frac{xy(x^2 - y^2)}{x^2 + y^2}$ on $[-1, 1] \times [-1, 1]$
- (c) $h(x, y) = y^2 - y^4 - x^2$ on $[-1, 1] \times [-1, 1]$
- (d) $k(x, y) = e^{-y} \cos x$ on $[-2\pi, 2\pi] \times [-2, 0]$

PROBLEM 1.5. Find the points of the hyperbolic cylinder $x^2 = z^2 - 1 = 0$ in \mathbb{R}^3 that are closest to the origin.

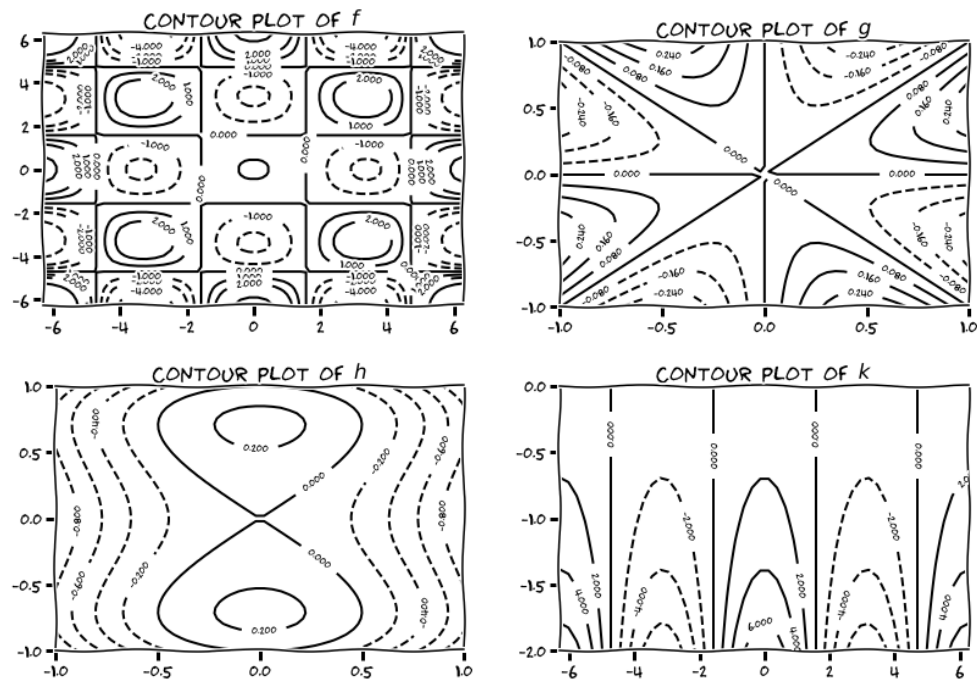


FIGURE 1.3. Contour plots for problem 1.4

CHAPTER 2

Existence of Extrema

In this chapter we will study different properties of functions and domains that guarantee existence of extrema. We start with a reminder of the definition of a continuous function.

DEFINITION. We say that a real-valued function $f: D \rightarrow \mathbb{R}$ is continuous at a point $\mathbf{x}_0 \in D$ if for all $\varepsilon > 0$ there exists $\delta > 0$ so that for all $\mathbf{x} \in D$ satisfying $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, it is $|f(\mathbf{x}) - f(\mathbf{x}_0)| < \varepsilon$.

EXAMPLE 2.1. Let $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ be given by

$$f(x, y) = \begin{cases} \frac{2xy}{x^2+y^2}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0) \end{cases}$$

This function is trivially continuous at any point $(x, y) \neq (0, 0)$. However, it fails to be continuous at the origin. Notice how we obtain different values as we approach $(0, 0)$ through different generic lines $y = mx$ with $m \in \mathbb{R}$:

$$\lim_{x \rightarrow 0} f(x, mx) = \lim_{x \rightarrow 0} \frac{2mx^2}{(1+m^2)x^2} = \frac{2m}{1+m^2}.$$

1. Continuous functions on compact domains

The existence of global maxima and minima is guaranteed for continuous functions over compact sets thanks to the following two basic results:

THEOREM 2.1 (Bounded Value Theorem). *The image $f(K)$ of a continuous real-valued function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ on a compact set K is bounded: there exists $M > 0$ so that $|f(\mathbf{x})| \leq M$ for all $\mathbf{x} \in K$.*

THEOREM 2.2 (Extreme Value Theorem). *A continuous real-valued function $f: K \rightarrow \mathbb{R}$ on a compact set $K \subset \mathbb{R}^d$ takes on minimal and maximal values on K .*

2. Continuous functions on unbounded domains

Extra restrictions must be applied to the behavior of f in this case, if we want to guarantee the existence of extrema. We consider first an obvious example based on Example 1.3.

DEFINITION (Coercive functions). A continuous real-valued function f is said to be *coercive* if for all $M > 0$ there exists $R = R(M) > 0$ so that $f(\mathbf{x}) \geq M$ if $\|\mathbf{x}\| \geq R$.

REMARK 2.1. This is equivalent to the limit condition

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} f(\mathbf{x}) = +\infty.$$

EXAMPLE 2.2. We saw in Example 1.3 how even-degree polynomials with positive leading coefficients are coercive, and how this helped guarantee the existence of a minimum.

We must be careful assessing coerciveness of polynomials in higher dimension. Consider for example $p_2(x, y) = x^2 - 2xy + y^2$. Note how $p_2(x, x) = 0$ for any $x \in \mathbb{R}$, which proves p_2 is not coercive.

To see that the polynomial $p_4(x, y) = x^4 + y^4 - 3xy$ is coercive, we start by factoring the leading terms:

$$x^4 + y^4 - 3xy = (x^4 + y^4) \left(1 - \frac{3xy}{x^4 + y^4} \right)$$

Assume $r > 1$ is large, and that $x^2 + y^2 = r^2$. We have then

$$\begin{aligned} x^4 + y^4 &\geq \frac{r^4}{2} & (\text{Why?}) \\ |xy| &\leq \frac{r^2}{2} & (\text{Why?}) \end{aligned}$$

therefore,

$$\begin{aligned} \frac{3xy}{x^4 + y^4} &\leq \frac{3}{r^2} \\ 1 - \frac{3xy}{x^4 + y^4} &\geq 1 - \frac{3}{r^2} \\ (x^4 + y^4) \left(1 - \frac{3xy}{x^4 + y^4} \right) &\geq \frac{r^2(r^2 - 3)}{2} \end{aligned}$$

We can then conclude that given $M > 0$, if $x^2 + y^2 \geq \frac{1}{2}(3 + \sqrt{9 + 8M})$, then $f(x, y) \geq M$.

THEOREM 2.3. *Coercive functions always have a global minimum.*

PROOF. Since f is coercive, there exists $r > 0$ so that $f(\mathbf{x}) > f(\mathbf{0})$ for all \mathbf{x} satisfying $\|\mathbf{x}\| > r$. On the other hand, consider the closed ball $K_r = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| \leq r\}$. The continuity of f guarantees a global minimum $\mathbf{x}^* \in K_r$ with $f(\mathbf{x}^*) \leq f(\mathbf{0})$. It is then $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$ trivially. \square

3. Convex functions

DEFINITION (Convex Sets). A subset $C \subseteq \mathbb{R}^d$ is said to be *convex* if for every $\mathbf{x}, \mathbf{y} \in C$, and every $\lambda \in [0, 1]$, the point $\lambda\mathbf{y} + (1 - \lambda)\mathbf{x}$ is also in C .

DEFINITION (Convex Functions). Given a convex set $C \subseteq \mathbb{R}^d$, we say that a real-valued function $f: C \rightarrow \mathbb{R}$ is *convex* if

$$f(\lambda\mathbf{y} + (1 - \lambda)\mathbf{x}) \leq \lambda f(\mathbf{y}) + (1 - \lambda)f(\mathbf{x})$$

If instead we have $f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})$ for $0 < \lambda < 1$, we say that the function f is said to be *concave* (resp. *strictly concave*) if $-f$ is convex (resp. strictly convex).

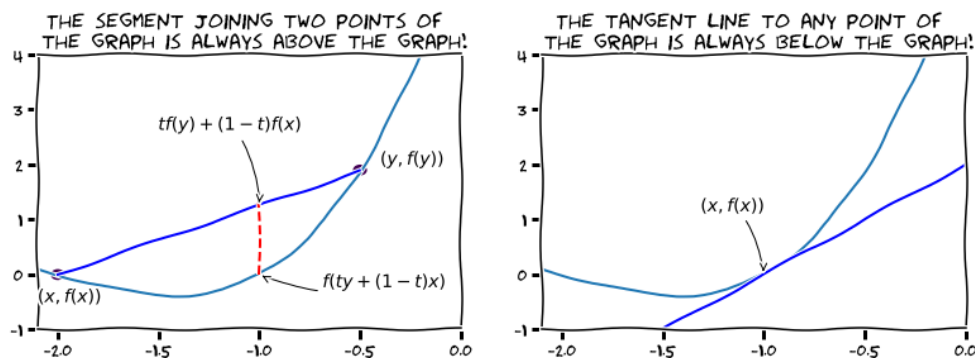


FIGURE 2.1. Convex Functions.

Exercises

PROBLEM 2.1. At what points $(x, y) \in \mathbb{R}^2$ is the function $f(x, y) = \frac{x + y}{2 + \cos x}$ continuous?

PROBLEM 2.2. Identify which of the following real-valued functions are coercive. Explain the reason.

- (a) $f(x, y) = \sqrt{x^2 + y^2}$.
- (b) $f(x, y) = x^2 + 9y^2 - 6xy$.
- (c) Rosenbrock functions $\mathcal{R}_{a,b}$.

PROBLEM 2.3. Find an example of a continuous, real-valued, non-coercive function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ that satisfies, for all $t \in \mathbb{R}$,

$$\lim_{x \rightarrow \infty} f(x, tx) = \lim_{y \rightarrow \infty} f(ty, y) = \infty$$

PROBLEM 2.4. Prove that convex functions $f: \mathbb{R}^d \rightarrow \mathbb{R}$ are continuous.

CHAPTER 3

Characterization of Extrema

CHAPTER 4

Nonlinear optimization

Bibliography

- [1] Ross L Finney, Maurice D Weir, and George Brinton Thomas. *Thomas' calculus: early transcendentals*. Addison-Wesley, 2001.
- [2] Anthony L Peressini, Francis E Sullivan, and J Jerry Uhl. *The mathematics of nonlinear programming*. Springer-Verlag New York, 1988.