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

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 $f\colon D \to \mathbb{R}$ 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 characterization of minima:

THEOREM 1.1. *Let $D \subseteq \mathbb{R}^2$ and let $f\colon D \to \mathbb{R}$ 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.

We employed the *Second Derivative Test for Local Extreme Values* to characterize some minima:

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

$$\frac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0 \tag{1}$$

$$\det \underbrace{\begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2}(x_0, y_0) & \dfrac{\partial^2 f}{\partial x \partial y}(x_0, y_0) \\[2ex] \dfrac{\partial^2 f}{\partial y \partial x}(x_0, y_0) & \dfrac{\partial^2 f}{\partial y^2}(x_0, y_0) \end{bmatrix}}_{\mathrm{Hess} f(x_0, y_0)} > 0 \tag{2}$$

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.

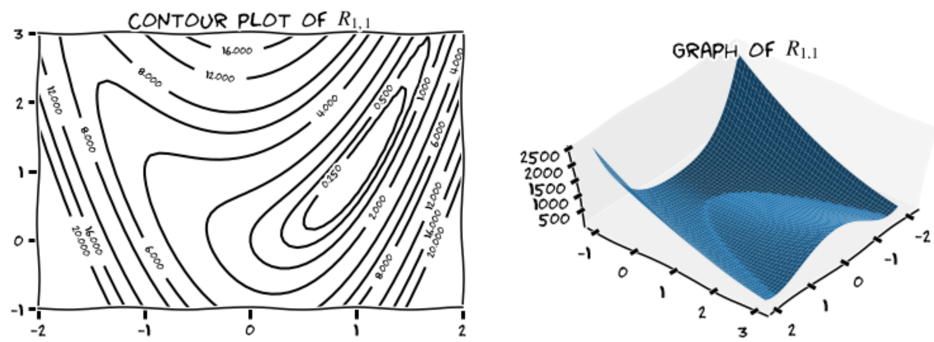

Figure 1.1. Details of the graph of $\mathcal{R}_{1,1}$

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

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

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

- 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:

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

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. If the function is also differentiable in the* interior *of 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:

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

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

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

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 $\lim_{|x|\to\infty} p_n(x) = +\infty$. To see this, we may write

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0 = a_n x^n \big(1 + \tfrac{a_{n-1}}{a_n x} + \cdots + \tfrac{a_0}{a_n x^n}\big)$$

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

$$\lim_{|x|\to\infty} a_n x^n = +\infty,$$

$$\lim_{|x|\to\infty} \left(1 + \tfrac{a_{n-1}}{a_n x} + \cdots + \tfrac{a_0}{a_n x^n}\right) = 1.$$

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, $\lim_{|x|\to\infty} f(x) = +\infty$, since $\lim_{|x|\to\infty} h(x) = +\infty$. 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 $x = \pm 1$.

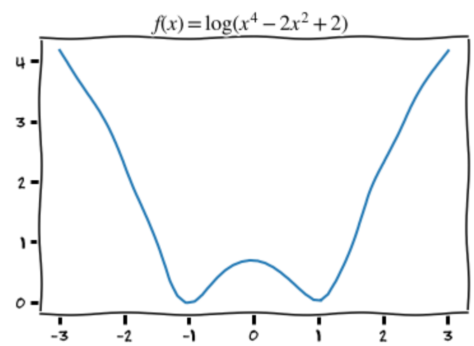

FIGURE 1.2. Global minima in unbounded domains

We have other useful characterizations for extrema, when the domain can be expressed as solutions of equations:

THEOREM 1.4 (Orthogonal Gradient). *Suppose $f(x,y)$ is differentiable in a region whose interior contains a smooth curve $C\colon \boldsymbol{r}(t) = \big(x(t), y(t)\big)$. If P_0 is a point on C where f has a local extremum relative to its values on C, then ∇f is orthogonal to C at P_0.*

This result leads to the *Method of Lagrange Multipliers*

THEOREM 1.5 (Lagrange Multipliers on one constraint). *Suppose that $f(x,y)$ and $g(x,y)$ are differentiable and $\nabla g \neq 0$ when $g(x,y,z) = 0$. To find the local extrema of f subject to the constraint $g(x,y) = 0$ (if these exist), find the values of x, y and λ that simultaneously satisfy the equations*

$$\nabla f = \lambda \nabla g, \text{ and } g(x,y) = 0$$

EXAMPLE 1.5. Find the minimum value of the expression $3x + 4y$ for values of x and y on the circle $x^2 + y^2 = 1$.

We start by modeling this problem to adapt the technique of Lagrange multipliers:

$$f(x,y) = \underbrace{3x+4y}_{\text{target}} \qquad g(x,y) = \underbrace{x^2+y^2-1}_{\text{constraint}}$$

Look for the values of x, y and λ that satisfy the equations $\nabla f = \lambda \nabla g$, $g(x,y) = 0$

$$3 = 2\lambda x, \qquad 4 = 2\lambda y \qquad 1 = x^2 + y^2$$

Equivalently, $\lambda \neq 0$ and x, y satisfy

$$x = \frac{3}{2\lambda}, \qquad y = \frac{2}{\lambda}, \qquad 1 = \frac{9}{4\lambda^2} + \frac{4}{\lambda^2}$$

These equations lead to $\lambda = \pm\frac{5}{2}$, and there are only two possible candidates for minimum. Evaluation of f on those gives that the minimum is at $\big(-\frac{3}{5}, -\frac{4}{5}\big)$.

This method can be extended to more than two dimensions, and more than one constraint. For instance:

Theorem 1.6 (Lagrange Multipliers on two constraints). *Suppose that $f(x,y,z)$, $g_1(x,y,z)$, $g_2(x,y,z)$ are differentiable with ∇g_1 not parallel to ∇g_2. To find the local extrema of f subject to the constraint $g_1(x,y,z) = g_2(x,y,z) = 0$ (if these exist), find the values of x, y, λ and μ that simultaneously satisfy the equations*

$$\nabla f = \lambda \nabla g_1 + \mu \nabla g_2, \qquad g_1(x,y,z) = 0, \qquad g_2(x,y,z) = 0$$

Example 1.6. The cylinder $x^2+y^2 = 1$ intersects the plane $x+y+z = 1$ in an ellipse. Find the points on the ellipse that lie closest to the origin.

We again model this as a Lagrange multipliers problem:

$$f(x,y,z) = \underbrace{x^2+y^2+z^2}_{\text{target}}, \quad g_1(x,y,z) = \underbrace{x^2+y^2-1}_{\text{constraint}}, \quad g_2(x,y,z) = \underbrace{x+y+z-1}_{\text{constraint}}.$$

The gradient equation $\nabla f = \lambda \nabla g_1 + \mu \nabla g_2$ gives

$$2x = 2\lambda x + \mu, \qquad 2y = 2\lambda y + \mu, \qquad 2z = \mu$$

These equations are satisfied simultaneously only in two scenarios:

(a) $\lambda = 1$ and $z = 0$
(b) $\lambda \neq 1$ and $x = y = z/(1-\lambda)$

Resolving each case we find four candidates: $(1,0,0)$, $(0,1,0)$, $(\sqrt{2}/2, \sqrt{2}/2, 1-\sqrt{2})$, and $(-\sqrt{2}/2, -\sqrt{2}/2, 1+\sqrt{2})$. The first two are our solution.

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

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

  For instance, the *open ball* of radius $r>0$ centered at a point $\boldsymbol{x}_0$ is the set $B_r(\boldsymbol{x}_0) = \{\boldsymbol{x}\in\mathbb{R}^d : \|\boldsymbol{x}-\boldsymbol{x}_0\| < r\}$.
- Given a real-valued function $f\colon D\to\mathbb{R}$ on a domain $D\subseteq\mathbb{R}^d$, we define the concept of *extrema*:

  Definition. Given $D\subseteq\mathbb{R}^d$, and a real-valued function $f\colon D\to\mathbb{R}$, we say that a point $\boldsymbol{x}^\star\in D$ is a:

  global minimum: $f(\boldsymbol{x}^\star)\le f(\boldsymbol{x})$ for all $\boldsymbol{x}\in D$.
  global maximum: $f(\boldsymbol{x}^\star)\ge f(\boldsymbol{x})$ for all $\boldsymbol{x}\in D$.
  strict global minimum: $f(\boldsymbol{x}^\star) < f(\boldsymbol{x})$ for all $\boldsymbol{x}\in D\setminus\{\boldsymbol{x}^\star\}$.
  strict global maximum: $f(\boldsymbol{x}^\star) > f(\boldsymbol{x})$ for all $\boldsymbol{x}\in D\setminus\{\boldsymbol{x}^\star\}$.
  local minimum: There exists $\delta>0$ so that $f(\boldsymbol{x}^\star)\le f(\boldsymbol{x})$ for all $\boldsymbol{x}\in B_\delta(\boldsymbol{x}^\star)\cap D$.
  local maximum: There exists $\delta>0$ so that $f(\boldsymbol{x}^\star)\ge f(\boldsymbol{x})$ for all $\boldsymbol{x}\in B_\delta(\boldsymbol{x}^\star)\cap D$.
  strict local minimum: There exists $\delta>0$ so that $f(\boldsymbol{x}^\star) < f(\boldsymbol{x})$ for all $\boldsymbol{x}\in B_\delta(\boldsymbol{x}^\star)\cap D$, $\boldsymbol{x}\ne\boldsymbol{x}^\star$.
  strict local maximum: There exists $\delta>0$ so that $f(\boldsymbol{x}^\star) > f(\boldsymbol{x})$ for all $\boldsymbol{x}\in B_\delta(\boldsymbol{x}^\star)\cap D$, $\boldsymbol{x}\ne\boldsymbol{x}^\star$.

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.
Characterization of extrema: Establish results that describe conditions for points $\boldsymbol{x}\in D$ to be extrema of f.
Tracking extrema: Design robust numerical algorithms that find the extrema for scientific computing purposes. This is the core of this course.

The development of existence and characterization results will be covered in chapter 2. The design of algorithms to track extrema will be covered in chapter 3.

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) $f(x,y) = \log\left(x^2 + y^2 - 4\right)$
(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) $g(x,y) = \dfrac{xy(x^2 - y^2)}{x^2 + y^2}$ 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 $[-2\pi, 2\pi] \times [-2, 0]$

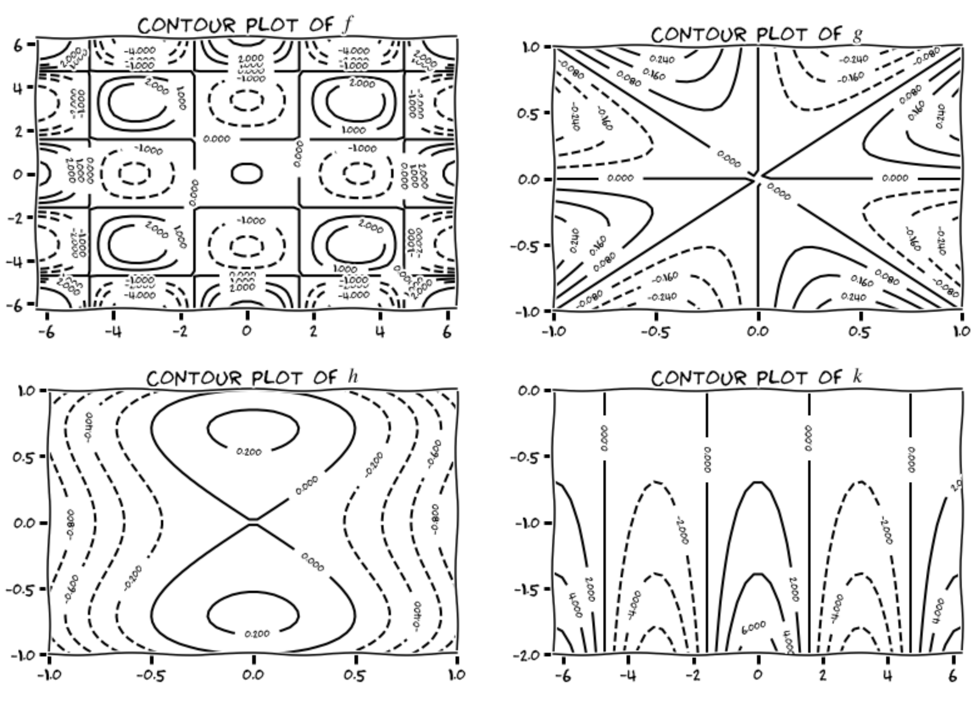

FIGURE 1.3. Contour plots for problem 1.4

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.

CHAPTER 2

Existence and Characterization of Extrema

In this chapter we will study different properties of functions and domains that guarantee existence of extrema. Once we have them, we explore characterization of those points. We start with a reminder of the definition of continuous and differentiable functions.

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

EXAMPLE 2.1. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ 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 $m \in \mathbb{R}$:

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

DEFINITION. A real-valued function f is said to be *differentiable* at $\boldsymbol{x}_0$ if there exists a *linear function* $J\colon \mathbb{R}^d \to \mathbb{R}$ so that

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{|f(\boldsymbol{x}_0 + h) - f(\boldsymbol{x}_0) - J(\boldsymbol{h})|}{\|\boldsymbol{h}\|} = 0$$

REMARK 2.1. A function is said to be *linear* if it satisfies $J(\boldsymbol{x} + \lambda\boldsymbol{y}) = J(\boldsymbol{x}) + \lambda J(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$. For each real-valued linear function $J\colon \mathbb{R}^d \to \mathbb{R}$ there exists $\boldsymbol{a} \in \mathbb{R}^d$ so that $J(\boldsymbol{x}) = \langle \boldsymbol{a}, \boldsymbol{x}\rangle$ for all $\boldsymbol{x} \in \mathbb{R}^d$. For this reason, the graph of a linear function is a hyperplane in $\mathbb{R}^d$.

REMARK 2.2. For any differentiable real-valued function f at a point $\boldsymbol{x}$ of its domain, the corresponding linear function in the definition above guarantees a tangent hyperplane to the graph of f at $\boldsymbol{x}$.

EXAMPLE 2.2. Consider a real-valued function $f\colon \mathbb{R} \to \mathbb{R}$ of a real variable. To prove differentiability at a point x_0, we need a linear function: $J(h) = ah$ for some $a \in \mathbb{R}$. Notice how in that case,

$$\frac{|f(x_0+h) - f(x_0) - J(h)|}{|h|} = \left| \frac{f(x_0) - f(x_0)}{h} - a \right|;$$

therefore, we could pick $a = \lim_{h\to 0} h^{-1}\big(f(x_0+h) - f(x_0)\big)$—this is the definition of derivative we learned in Calculus: $a = f'(x_0)$

A *friendly* version of the differentiability of real-valued functions comes with the next result (see, e.g. [**1**, p.818])

THEOREM 2.1. *If the partial derivatives* $\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_d}$ *of a real-valued function* $f\colon \mathbb{R}^d \to \mathbb{R}$ *are continuous on an open region* $G \subseteq \mathbb{R}^d$, *then* f *is differentiable at every point of* $\mathbb{R}$.

EXAMPLE 2.3. Let $f\colon \mathbb{R}^d \to \mathbb{R}$. To prove that f is differentiable at a point $\boldsymbol{x}_0 \in \mathbb{R}^d$ we need a linear function $J(h) = \langle \boldsymbol{a}, h\rangle$ for some $\boldsymbol{a} \in \mathbb{R}^d$. Under the conditions of Theorem 2.1 we may use

$$\boldsymbol{a} = \nabla f(\boldsymbol{x}_0) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0), \dots, \frac{\partial f}{\partial x_d}(\boldsymbol{x}_0) \right).$$

It is a simple task to prove that all differentiable functions are continuous. Is it true that all continuous functions are differentiable?

EXAMPLE 2.4 (Weierstrass Function). For any positive real numbers a, b satisfying $0 < a < 1 < b$ and $ab \geq 1$, consider the Weierstrass function $\mathcal{W}_{a,b}\colon \mathbb{R} \to \mathbb{R}$ given by

$$\mathcal{W}_{a,b}(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$

This function is continuous everywhere, yet *nowhere* differentiable! For a proof, see e.g. [**2**]

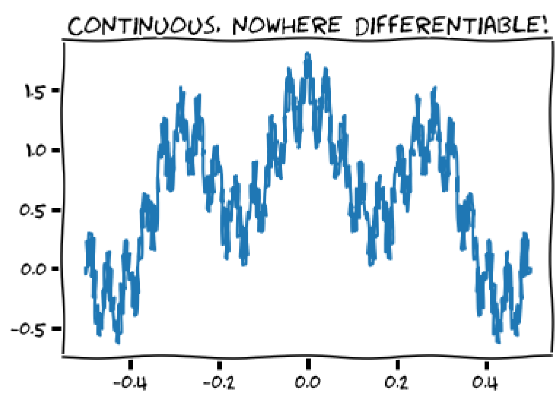

FIGURE 2.1. Detail of the graph of $\mathcal{W}_{0.5,7}$

1. Existence

1.1. Continuous functions on compact domains. The existence of global extrema is guaranteed for continuous functions over compact sets thanks to the following two basic results:

THEOREM 2.2 (Bounded Value Theorem). *The image $f(K)$ of a continuous real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$ on a compact set K is bounded: there exists $M > 0$ so that $|f(\boldsymbol{x})| \le M$ for all $\boldsymbol{x} \in K$.*

THEOREM 2.3 (Extreme Value Theorem). *A continuous real-valued function $f\colon K \to \mathbb{R}$ on a compact set $K \subset \mathbb{R}^d$ takes on minimal and maximal values on K.*

1.2. 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 $f(\boldsymbol{x}) \ge M$ if $\|\boldsymbol{x}\| \ge R$.

REMARK 2.3. This is equivalent to the limit condition

$$\lim_{\|\boldsymbol{x}\|\to\infty} f(\boldsymbol{x}) = +\infty.$$

EXAMPLE 2.5. 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

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

therefore,

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

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

THEOREM 2.4. *Coercive functions always have a global minimum.*

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

2. Characterization

2.1. Differentiability and Characterization. Differentiability is key to guarantee characterization of extrema. Critical points lead the way:

THEOREM 2.5 (First order necessary optimality condition for minimization). *Suppose* $f\colon \mathbb{R}^d \to \mathbb{R}$ *is differentiable at* $\boldsymbol{x}^\star$*. If* $\boldsymbol{x}^\star$ *is a local minimum, then* $\nabla(f)(\boldsymbol{x}^\star) = 0$.

To be able to classify extrema in a properly differentiable function, it is necessary to see the behavior of the function with respect to the tangent hyperplane at our candidates. Second derivatives make this process very easy. Before we proceed to the key results, we should refresh some terminology:

DEFINITION (Hessian). Given a twice-differentiable function $f\colon D \subseteq \mathbb{R}^d \to \mathbb{R}$, we define the *Hessian* of f at $\boldsymbol{x} \in D$ to be the following matrix of second partial derivatives:

$$\operatorname{Hess} f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_d}(\boldsymbol{x}) \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \dfrac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_d}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_d \partial x_1}(\boldsymbol{x}) & \dfrac{\partial^2 f}{\partial x_d \partial x_2}(\boldsymbol{x}) & \cdots & \dfrac{\partial^2 f}{\partial x_d^2}(\boldsymbol{x}) \end{bmatrix}$$

THEOREM 2.6 (Second order necessary optimality condition for minimization). *Suppose that* $f\colon \mathbb{R}^d \to \mathbb{R}$ *is twice continuously differentiable at* $\boldsymbol{x}^\star$*. If* $\boldsymbol{x}^\star$ *is a local minimum, then* $\nabla f(\boldsymbol{x}^\star) = 0$ *and* $\operatorname{Hess} f(\boldsymbol{x}^\star)$ *is* positive semidefinite.

THEOREM 2.7 (Second order sufficient optimality condition for minimization). *Suppose* $f\colon \mathbb{R}^d \to \mathbb{R}$ *is twice differentiable at* $\boldsymbol{x}^\star$*. If* $\nabla f(\boldsymbol{x}^\star) = 0$ *and* $\operatorname{Hess} f(\boldsymbol{x}^\star)$ *is* positive definite*, then* $\boldsymbol{x}^\star$ *is a strict local minimum.*

2.2. Convex functions and Characterization.

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

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

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

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

Convex functions have many pleasant properties:

Theorem 2.8. *Convex functions are continuous*

Theorem 2.9. *Let $f : C \to \mathbb{R}$ be a real-valued convex function defined on a convex set $C \subseteq \mathbb{R}^d$. If $\lambda_1, \dots, \lambda_n$ are nonnegative numbers satisfying $\lambda_1 + \dots + \lambda_n = 1$ and $\boldsymbol{x}_1, \dots, x_n$ are n different points in C, then*

$$f\big(\lambda_1 \boldsymbol{x}_1 + \dots + \lambda_n x_n\big) \leq \lambda_1 f(\boldsymbol{x}_1) + \dots + \lambda_n f(\boldsymbol{x}_n).$$

Theorem 2.10. *If $f : C \to \mathbb{R}$ is a function on a convex set $C \subseteq \mathbb{R}^d$ with continuous first partial derivatives on C, then*

(a) *f is convex if and only if for all $\boldsymbol{x}, \boldsymbol{y} \in C$,*

$$f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \leq f(\boldsymbol{y}).$$

(b) *f is strictly convex if for all $\boldsymbol{x} \neq \boldsymbol{y} \in C$,*

$$f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle < f(\boldsymbol{y}).$$

Remark 2.4. Theorem 2.10 implies that the graph of any (strictly) convex function always lies over the tangent hyperplane at any point of the graph.

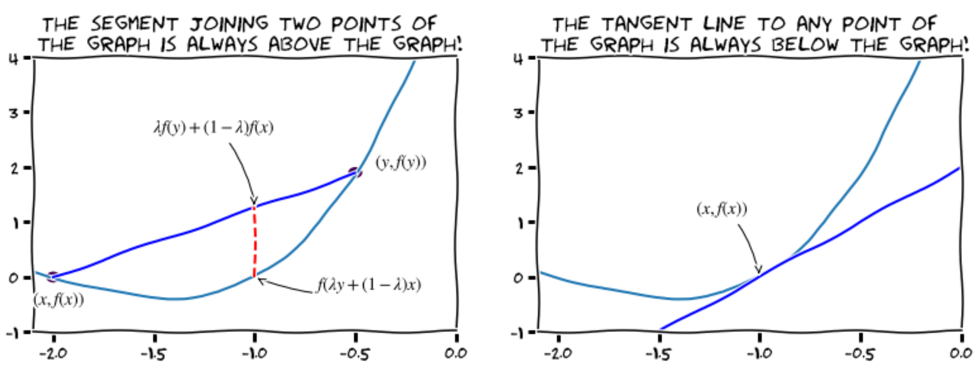

Figure 2.2. Convex Functions.

Another useful characterization of convex functions.

Theorem 2.11. *Suppose that $f : C \to \mathbb{R}$ is a function with second partial derivatives on an open convex set $C \subseteq \mathbb{R}^d$. If the Hessian is positive semidefinite (resp. positive definite) on C, then f is convex (resp. strictly convex).*

Exercises

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

PROBLEM 2.2. For the following optimization problems, state whether existence of a solution is guaranteed:

(a) $f(x) = \frac{1+x}{2x}$ over $[1, \infty)$
(b) $f(x) = 1/x$ over $[1, 2)$
(c) The pieewise function $f(x)$ below over $[1, 2]$

$$f(x) = \begin{cases} 1/x, & x < 2 \\ 1, & x = 2 \end{cases}$$

PROBLEM 2.3. 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 2.4. Find an example of a continuous, real-valued, non-coercive function $f\colon \mathbb{R}^2 \to \mathbb{R}$ that satisfies, for all $t \in \mathbb{R}$,

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

CHAPTER 3

Nonlinear optimization

Bibliography

[1] Ross L Finney, Maurice D Weir, and George Brinton Thomas. *Thomas' calculus: early transcendentals*. Addison-Wesley, 2001.

[2] Godefroy Harold Hardy. Weierstrasss non-differentiable function. *Trans. Amer. Math. Soc*, 17(3):301–325, 1916.

[3] Anthony L Peressini, Francis E Sullivan, and J Jerry Uhl. *The mathematics of nonlinear programming*. Springer-Verlag New York, 1988.