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. [**3**, Chapter 14])

DEFINITION. If $f(x, y)$ is differentiable in an open region containing the point (x_0, y_0), we define the *gradient vector* of $f(x, y)$ at (x_0, y_0) as the vector

$$\nabla f(x_0, y_0) = \left[\frac{\partial f(x_0, y_0)}{\partial x}, \frac{\partial f(x_0, y_0)}{\partial y}\right].$$

Given any vector $\boldsymbol{v} = [v_1, v_2]$ with $\|\boldsymbol{v}\| = \left(v_1^2 + v_2^2\right)^{1/2} = 1$ (what we call a *unit vector* or a *direction*), we define the *directional derivative* of f in the direction $\boldsymbol{v}$ at (x_0, y_0) by

$$D_{\boldsymbol{v}} f(x_0, y_0) = \langle \nabla f(x_0, y_0), \boldsymbol{v} \rangle = v_1 \frac{\partial f(x_0, y_0)}{\partial x} + v_2 \frac{\partial f(x_0, y_0)}{\partial y}.$$

REMARK 1.1. The gradient has many interesting properties. Assume $f(x, y)$ is a differentiable function.

Fastest Increase: At any point (x, y), the function f increases most rapidly in the direction of the gradient vector $\boldsymbol{v} = \nabla f(x, y)$. The derivative in that direction is $D_{\boldsymbol{v}} f(x, y) = \|\nabla f(x, y)\|$.

Fastest Decrease: At any point (x, y), the function f decreases most rapidly in the direction $\boldsymbol{v} = -\nabla f(x, y)$. The derivative in that direction is $D_{\boldsymbol{v}} f(x, y) = -\|\nabla f(x, y)\|$.

Zero Change: Any direction $\boldsymbol{v}$ perpendicular to a non-zero gradient is a direction of *zero change* in f at (x, y): $D_{\boldsymbol{v}} f(x, y) = 0$.

Tangents to Level Curves: At every point (x, y) in the domain of f, the gradient $\nabla f(x, y)$ is perpendicular to the level curve through (x, y).

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:*

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

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

REMARK 1.2. 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 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.

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

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

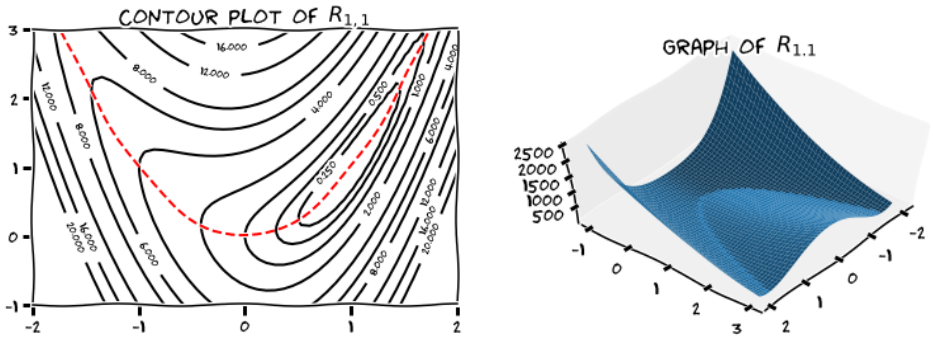

FIGURE 1.1. Details of the graph of $\mathcal{R}_{1,1}$

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

- 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), which happens also to be a global minimum.

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:

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

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

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

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 $x = \pm 1$.

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

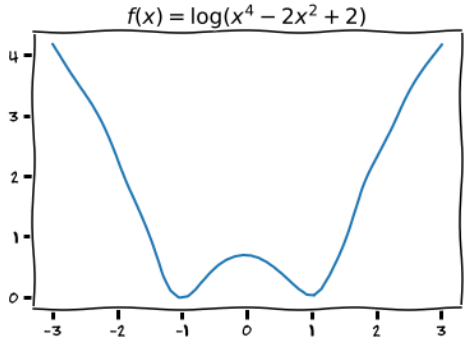

FIGURE 1.2. Global minima in unbounded domains

THEOREM 1.4 (Orthogonal Gradient). *Suppose $f(x, y)$ is differentiable in a region whose interior contains a smooth curve C: $\boldsymbol{r}(t) = \big(x(t), y(t)\big)$. 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 $g(x, y) = 0$. 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 the point $\big(-\frac{3}{5}, -\frac{4}{5}\big)$.

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:

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

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:

$$(1,0,0), \quad (0,1,0), \quad (\sqrt{2}/2, \sqrt{2}/2, 1-\sqrt{2}), \quad (-\sqrt{2}/2, -\sqrt{2}/2, 1+\sqrt{2}).$$

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

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

  For instance, the *open ball* of radius $r > 0$ centered at a point $\boldsymbol{x}^\star$ is the set $B_r(\boldsymbol{x}^\star) = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{x}^\star\| < r\}$.
- 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* and *extreme Values*:

  DEFINITION. Given a real-valued function $f\colon D \to \mathbb{R}$ on a domain $D \subseteq \mathbb{R}^d$, we say that a point $\boldsymbol{x}^\star \in D$ is a:
  global minimum: $f(\boldsymbol{x}^\star) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in D$.
  global maximum: $f(\boldsymbol{x}^\star) \geq f(\boldsymbol{x})$ 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 $f(\boldsymbol{x}^\star) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in B_\delta(\boldsymbol{x}^\star) \cap D$.
local maximum: There exists $\delta > 0$ so that $f(\boldsymbol{x}^\star) \geq f(\boldsymbol{x})$ 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 $\boldsymbol{x} \in B_\delta(\boldsymbol{x}^\star) \cap D$, $\boldsymbol{x} \neq \boldsymbol{x}^\star$.
strict local maximum: There exists $\delta > 0$ so that $f(\boldsymbol{x}^\star) > f(\boldsymbol{x})$ for all $\boldsymbol{x} \in B_\delta(\boldsymbol{x}^\star) \cap D$, $\boldsymbol{x} \neq \boldsymbol{x}^\star$.

In this setting, the objective of *optimization* is the search for extrema in the following two scenarios:

Unconstrained Optimization: if D is an open set (usually the whole space $\mathbb{R}^d$).
Constrained Optimization: if D can be described as a set of *constraints*: $\boldsymbol{x} \in D$ if there exist $m, n \in \mathbb{N}$ and functions $g_k \colon \mathbb{R}^d \to \mathbb{R}$ $(1 \leq k \leq m)$, $h_j \colon \mathbb{R}^d \to \mathbb{R}$ $(1 \leq j \leq n)$ so that

$$g_k(\boldsymbol{x}) \leq 0 \quad (1 \leq k \leq m)$$
$$h_j(\boldsymbol{x}) = 0 \quad (1 \leq j \leq n)$$

For each of these problems, we follow a similar 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.

The development of existence and characterization results for unconstrained optimization will be covered in chapter 2. The design of algorithms to track extrema in the unconstrained setting will be covered in chapter 3. Chapter 4 is devoted to existence and characterization results for constrained optimization, and Chapter 5 for the design of algorithms in that setting.

Exercises

PROBLEM 1.1 (Advanced). State and prove similar statements as in Definition 1, Theorems 1.1, 1.2 and 1.3, but for *local* and *global maxima*.

PROBLEM 1.2 (Basic). 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 (Basic). 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 (CAS). 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 $[-2\pi, 2\pi] \times [-2, 0]$

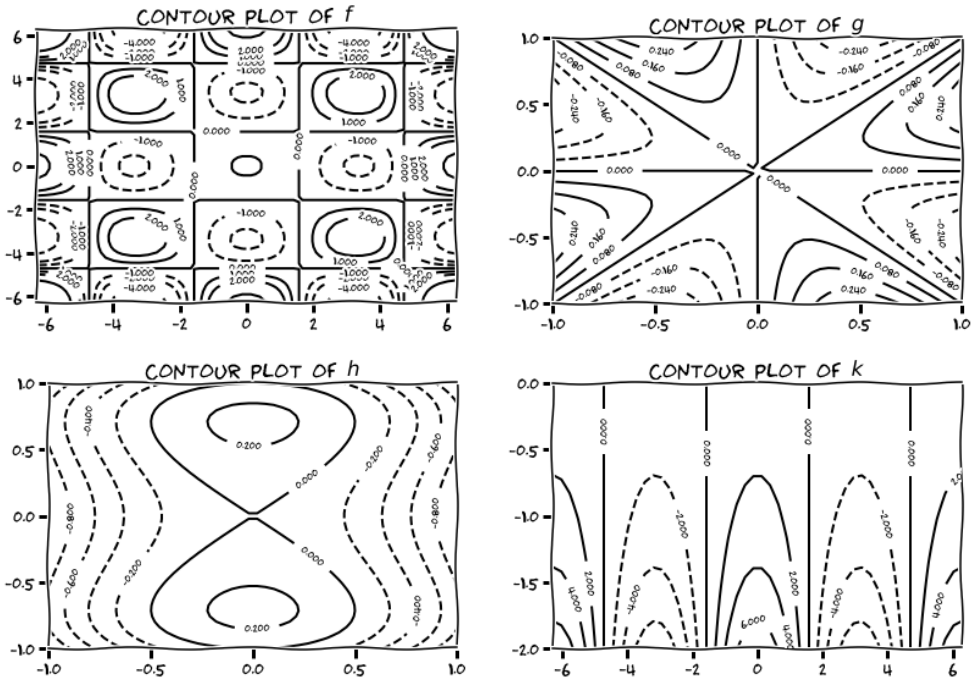

FIGURE 1.3. Contour plots for problem 1.4

PROBLEM 1.5 (Basic). Sketch the curve $f(x, y) = c$ together with ∇f and the tangent line at the given point. Write an equation for the tangent line.

(a) $f(x, y) = x^2 + y^2$, $c = 4$, $(\sqrt{2}, \sqrt{2})$.
(b) $f(x, y) = x^2 - y$, $c = 1$, $(\sqrt{2}, 1)$.
(c) $f(x, y) = xy$, $c = -1$, $(2, -2)$.
(d) $f(x, y) = x^2 - xy + y^2$, $c = 7$, $(-1, 2)$.

PROBLEM 1.6 (Basic). For the function

$$f(x,y) = \frac{x-y}{x+y},$$

at the point $P_0 = (-1/2, 3/2)$, find the directions $\boldsymbol{v}$ and the directional derivatives $D_{\boldsymbol{v}}f(P_0)$ for which

(a) $D_{\boldsymbol{v}}f(P_0)$ is largest.
(b) $D_{\boldsymbol{v}}f(P_0)$ is smallest.
(c) $D_{\boldsymbol{v}}f(P_0) = 0$.
(d) $D_{\boldsymbol{v}}f(P_0) = 1$.
(e) $D_{\boldsymbol{v}}f(P_0) = -2$.

PROBLEM 1.7 (Intermediate). The derivative of $f(x,y)$ at $(1,2)$ in the direction $\frac{\sqrt{2}}{2}[1,1]$ is $2\sqrt{2}$ and in the direction $[0,-1]$ is -3. What is the derivative of f in the direction $\frac{\sqrt{5}}{5}[-1,-2]$?

PROBLEM 1.8 (Intermediate). Find the absolute maxima and minima of the function $f(x,y) = (4x - x^2)\cos y$ on the rectangular plate $1 \le x \le 3, -\frac{\pi}{4} \le y \le \frac{\pi}{4}$.

PROBLEM 1.9 (Basic). Find two numbers $a \le b$ such that

$$\int_a^b (24 - 2x - x^2)^{1/3}\, dx$$

has its largest value.

PROBLEM 1.10 (Basic). Find the points of the hyperbolic cylinder $x^2 - z^2 - 1 = 0$ in $\mathbb{R}^3$ that are closest to the origin.

PROBLEM 1.11 (Intermediate). Find the extreme values of the function $f(x,y,z) = xy + z^2$ on the circle in which the plane $y - x = 0$ intersects the sphere $x^2 + y^2 + z^2 = 4$.

PROBLEM 1.12 (CAS). Write a routine (in your favorite CAS) that uses *symbolic computation*[1] to find the minimum of a differentiable real-valued function $f\colon \mathbb{R} \to \mathbb{R}$ over

(a) a closed interval $[a,b]$
(b) An interval of the form $[a,\infty)$, or $(-\infty,b]$

The routine should accept as input:

- the expression of the function f,
- the endpoints a, b.

[1] See Appendix A

CHAPTER 2

Existence and Characterization of Extrema for Unconstrained Optimization

In this chapter we will study different properties of functions and domains that guarantee existence of extrema for unconstrained optimization. Once we have them, we explore characterization of those points. We start with a reminder of the definition of continuous and differentiable functions, and then we proceed to introduce other functions with advantageous properties for optimization purposes.

1. Functions

1.1. Continuity and Differentiability.

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

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 $\boldsymbol{x}^\star$ if there exists a *linear function* $J\colon \mathbb{R}^d \to \mathbb{R}$ so that

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

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 $x^\star$, we need a linear function: $J(h) = ah$ for some $a \in \mathbb{R}$. Notice how in that case,

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

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

A *friendly* version of the differentiability of real-valued functions comes with the next result (see, e.g. [**3**, p.818])

THEOREM 2.1. *If the partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_d}$ 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 $\boldsymbol{x}^\star \in \mathbb{R}^d$ 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

$$\boldsymbol{a} = \nabla f(\boldsymbol{x}^\star) = \left[\frac{\partial f(\boldsymbol{x}^\star)}{\partial x_1}, \ldots, \frac{\partial f(\boldsymbol{x}^\star)}{\partial x_d}\right].$$

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 *nowehere* differentiable! (see Figure 2.1). For a proof, see e.g. [**6**]

A few more useful results about higher order derivatives follow:

THEOREM 2.2 (Clairaut). *If $f\colon \mathbb{R}^d \to \mathbb{R}$ and its partial derivatives of orders 1 and 2, $\frac{\partial f}{\partial x_k}$, $\frac{\partial^2 f}{\partial x_k \partial x_j}$, $(1 \leq k, j \leq d)$ are defined throughout an open region containing the point $\boldsymbol{x}^\star$, and are all continuous at $\boldsymbol{x}^\star$, then*

$$\frac{\partial^2 f(\boldsymbol{x}^\star)}{\partial x_k \partial x_j} = \frac{\partial^2 f(\boldsymbol{x}^\star)}{\partial x_j \partial x_k}, \quad (1 \leq k, j \leq d).$$

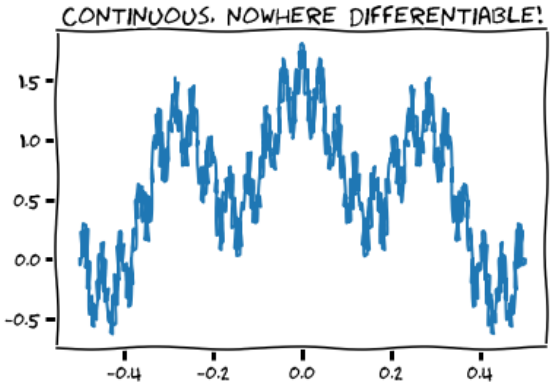

FIGURE 2.1. Detail of the graph of $\mathcal{W}_{0.5,7}$

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

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

Functions that satisfy the conditions of Theorem 2.2 have symmetric Hessians. We shall need some properties in regard to symmetric matrices.

DEFINITION. Given a symmetric matrix $\boldsymbol{A}$, we define its associated *quadratic form* as the function $\mathcal{Q}_{\boldsymbol{A}}\colon \mathbb{R}^d \to \mathbb{R}$ given by

$$\mathcal{Q}_{\boldsymbol{A}}(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{A}\boldsymbol{x}^{\intercal} = \begin{bmatrix} x_1 \cdots x_d \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{1d} & \cdots & a_{dd} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

We say that a symmetric matrix is:

positive definite: if $\mathcal{Q}_{\boldsymbol{A}}(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$.
positive semidefinite: if $\mathcal{Q}_{\boldsymbol{A}}(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^d$.
negative definite: if $\mathcal{Q}_{\boldsymbol{A}}(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$.
negative semidefinite: if $\mathcal{Q}_{\boldsymbol{A}}(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x} \in \mathbb{R}^d$.
indefinite: if there exist $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ so that $\mathcal{Q}_{\boldsymbol{A}}(\boldsymbol{x})\mathcal{Q}_{\boldsymbol{A}}(\boldsymbol{y}) < 0$.

EXAMPLE 2.5. Let $\boldsymbol{A}$ be the 3×3–symmetric matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 2 \\ -1 & 3 & 0 \\ 2 & 0 & 5 \end{bmatrix}$$

The associated quadratic form is given by

$$\begin{aligned} \mathcal{Q}_{\boldsymbol{A}}(x,y,z) &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 2 & -1 & 2 \\ -1 & 3 & 0 \\ 2 & 0 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 2x - y + 2z \\ -x + 3y \\ 2x + 5z \end{bmatrix} \\ &= x(2x - y + 2z) + y(-x + 3y) + z(2x + 5z) \\ &= 2x^2 + 3y^2 + 5z^2 - 2xy + 4xz \end{aligned}$$

To easily classify symmetric matrices, we usually employ any of the following two criteria:

THEOREM 2.3 (Principal Minor Criteria). *Given a general square matrix $\boldsymbol{A}$, we define for each $1 \leq \ell \leq d$, Δ_ℓ (the ℓth* principal minor *of $\boldsymbol{A}$) to be the determinant of the upper left-hand corner $\ell \times \ell$–submatrix of $\boldsymbol{A}$.*

$$\begin{array}{c} \begin{array}{ccccc} \Delta_1 & \Delta_2 & \Delta_3 & & \end{array} \\ \left[\begin{array}{c|c|c|ccc} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \hline a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \hline a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{array}\right] \end{array}$$

A symmetric matrix $\boldsymbol{A}$ is:

- *Positive definite if and only if $\Delta_\ell > 0$ for all $1 \leq \ell \leq d$.*
- *Negative definite if and only if $(-1)^\ell \Delta_\ell > 0$ for all $1 \leq \ell \leq d$.*

THEOREM 2.4 (Eigenvalue Criteria). *Given a general square $d \times d$ matrix $\boldsymbol{A}$, consider the function $p_{\boldsymbol{A}} \colon \mathbb{C} \to \mathbb{C}$ given by $p_{\boldsymbol{A}}(\lambda) = \det\big(\boldsymbol{A} - \lambda \boldsymbol{I}_d\big)$. This is a polynomial of (at most) degree d in λ. We call it the* characteristic polynomial *of $\boldsymbol{A}$. The roots (in $\mathbb{C}$) of the characteristic polynomial are called the* eigenvalues *of $\boldsymbol{A}$. Symmetric matrices enjoy the following properties:*

(a) The eigenvalues of a symmetric matrix are all real.
(b) If $\lambda \in \mathbb{R}$ is a root of multiplicity n of the characteristic polynomial of a (non-trivial) symmetric matrix, then there exist n linearly independent vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ satisfying $\boldsymbol{A}\boldsymbol{x}_k = \lambda \boldsymbol{x}_k$ $(1 \leq k \leq n)$.

(c) *If $\lambda_1 \neq \lambda_2$ are different roots of the characteristic polynomial of a symmetric matrix, and $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$ satisfy $\boldsymbol{A}\boldsymbol{x}_k = \lambda_k \boldsymbol{x}_k$ ($k = 1, 2$), then $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = 0$.*
(d) *A symmetric matrix is positive definite (resp. negative definite) if and only if all its eigenvalues are positive (resp. negative).*
(e) *A symmetric matrix is positive semidefinite (resp. negative semidefinite) if and only if all its eigenvalues are non-negative (resp. non-positive)*
(f) *A symmetric matrix is indefinite if there exist two eigenvalues $\lambda_1 \neq \lambda_2$ with different sign.*

1.2. Coercive Functions. Other set of functions that play an important role in optimization are the kind of functions we explored in 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 $f(\boldsymbol{x}) \geq M$ if $\|\boldsymbol{x}\| \geq R$.

REMARK 2.3. This is equivalent to the limit condition

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

EXAMPLE 2.6. 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 $p_4(x, y) = x^4 + y^4 - 4xy$ is coercive, we start by factoring the leading terms:

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

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

$$x^4 + y^4 \geq \frac{r^4}{2} \qquad \text{(Why?)}$$

$$|xy| \leq \frac{r^2}{2} \qquad \text{(Why?)}$$

therefore,

$$\frac{4xy}{x^4 + y^4} \leq \frac{4}{r^2}$$

$$1 - \frac{4xy}{x^4 + y^4} \geq 1 - \frac{4}{r^2}$$

$$\left(x^4 + y^4\right)\left(1 - \frac{4xy}{x^4 + y^4}\right) \geq \frac{r^2(r^2 - 4)}{2}$$

We can then conclude that given $M > 0$, if $x^2 + y^2 \geq 2 + \sqrt{4 + 2M}$, then $p_4(x, y) \geq M$. This proves p_4 is coercive.

1.3. Convex Functions. There is one more kind of functions we should explore.

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.

The following result is an interesting characterization of convex sets that allows us to actually construct any convex set from a family of points.

THEOREM 2.5. *Let $C \subseteq \mathbb{R}^d$ be a convex set and let $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \subset C$ be a family of points in C. The convex combinations $\lambda_1\boldsymbol{x}_1 + \lambda_2\boldsymbol{x}_2 + \cdots + \lambda_n\boldsymbol{x}_n$ are also in C, provided $\lambda_k \geq 0$ for all $1 \leq k \leq n$ and $\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1$.*

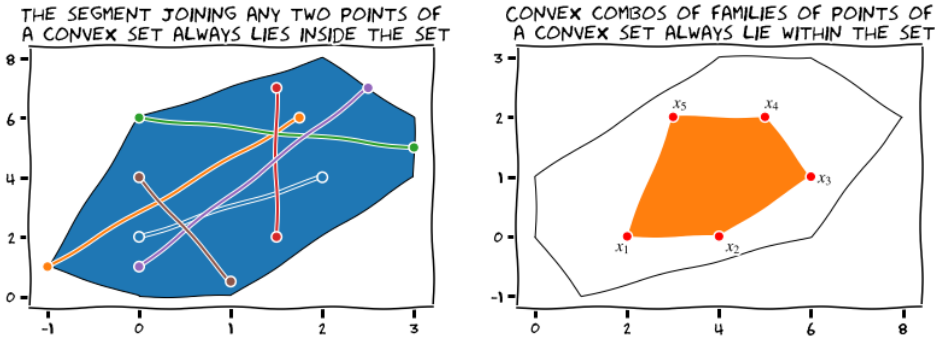

FIGURE 2.2. Convex sets.

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

REMARK 2.4. There is an alternative definition of convex functions using the concept of *epigraph* of a function. Given a convex function $f\colon C \to \mathbb{R}$ on a convex set C, the epigraph of f is a set $\operatorname{epi}(f) \subset \mathbb{R}^{d+1}$ defined by

$$\operatorname{epi}(f) = \{(\boldsymbol{x}, y) \in \mathbb{R}^{d+1} : \boldsymbol{x} \in C, y \in \mathbb{R}, f(\boldsymbol{x}) \leq y\}.$$

The function f is convex if and only if its epigraph is a convex set.

Convex functions have many pleasant properties:

THEOREM 2.6. *Convex functions are continuous.*

THEOREM 2.7. *Let $f\colon C \to \mathbb{R}$ 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 $\lambda_1 + \cdots + \lambda_n = 1$ and $\boldsymbol{x}_1, \dots, x_n$ are n different points in C, then*

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

THEOREM 2.8. *If $f\colon C \to \mathbb{R}$ 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 2.5. Theorem 2.8 implies that the graph of any (strictly) convex function always lies over the tangent hyperplane at any point of the graph.

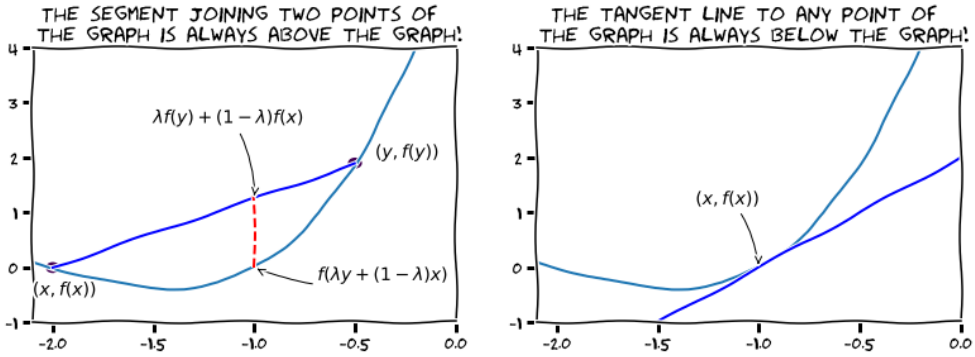

FIGURE 2.3. Convex Functions.

Two more useful characterization of convex functions.

THEOREM 2.9. *Suppose that $f\colon C \to \mathbb{R}$ 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).*

THEOREM 2.10. *Let $C \subseteq \mathbb{R}^d$ be a convex set.*

(a) If $f_k\colon C \to \mathbb{R}$ are convex functions for $1 \leq k \leq n$, then so is the sum $f\colon C \to \mathbb{R}$:

$$f(\boldsymbol{x}) = \sum_{k=1}^{n} f_k(\boldsymbol{x}).$$

If at least one of them is strictly convex, then so is f.

(b) *If $f\colon C \to \mathbb{R}$ is convex (resp. strictly convex) on C, then so is λf for any $\lambda > 0$.*
(c) *If $f\colon C \to \mathbb{R}$ is convex (resp. strictly convex) on C, and $g\colon f(C) \to \mathbb{R}$ is an increasing convex function (resp. strictly increasing convex), then so is $g \circ f$.*
(d) *If $f, g\colon C \to \mathbb{R}$ are convex functions on C, then so is* $\max\{f, g\}$.

EXAMPLE 2.7. Consider the function $f(x, y, z)$ defined on $\mathbb{R}^3$ by

$$f(x, y, z) = 2x^2 + y^2 + z^2 + 2yz.$$

Notice that for all $(x, y, z) \in \mathbb{R}^3$,

$$\text{Hess}f(x, y, z) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}, \quad \Delta_1 = 4 > 0, \quad \Delta_2 = 8 > 0, \quad \Delta_3 = 0.$$

By virtue of Theorem 2.9, we infer that the function f is convex, but not strictly convex.

EXAMPLE 2.8. To prove that $f(x, y, z) = e^{x^2+y^2+z^2}$ is convex, rather than computing the Hessian and address if it is positive (semi)definite, it is easier to realize that we can write $f = g \circ h$ with

$$\begin{aligned} &g\colon \mathbb{R} \to \mathbb{R} && h\colon \mathbb{R}^3 \to \mathbb{R} \\ &g(x) = e^x && h(x, y, z) = x^2 + y^2 + z^2 \end{aligned}$$

The function g is trivially strictly increasing and convex (since $g'(x) = g''(x) = e^x > 0$ for all $x \in \mathbb{R}$). The function h is strictly convex, since (by Theorem 2.9)

$$\text{Hess}h(x, y, z) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \Delta_1 = 2 > 0, \quad \Delta_2 = 4 > 0, \quad \Delta_3 = 8 > 0.$$

By virtue of (c) in Theorem 2.10, we infer that f is strictly convex.

EXAMPLE 2.9. Set $C = \{(x, y) \in \mathbb{R}^2 : x > 0, y > 0\}$. Consider the function $f\colon C \to \mathbb{R}$ given by

$$f(x, y) = x^2 - 4xy + 5y^2 - \log(xy)$$

Notice we may write $f = g + h$ with $g, h\colon C \to \mathbb{R}$ given respectively by $g(x, y) = x^2 - 4xy + 5y^2$ and $h(x, y) = -\log(xy)$. Note also that both functions are strictly convex, since for all $(x, y) \in C$:

$$\begin{aligned} &\text{Hess}g(x, y) = \begin{bmatrix} 2 & -4 \\ -4 & 10 \end{bmatrix}, && \Delta_1 = 2 > 0, && \Delta_2 = 4 > 0, \\ &\text{Hess}h(x, y) = \begin{bmatrix} x^{-2} & 0 \\ 0 & y^{-2} \end{bmatrix}, && \Delta_1 = x^{-2} > 0, && \Delta_2 = (xy)^{-2} > 0. \end{aligned}$$

By virtue of part (a) in Theorem 2.10, we infer that f is strictly convex.

We are now ready to explore existence and characterization of extrema in a wide variety of situations.

2. Existence

2.1. 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 2.11 (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 $|f(\boldsymbol{x})| \leq M$ for all $\boldsymbol{x} \in K$.*

THEOREM 2.12 (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.*

2.2. 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.

THEOREM 2.13. *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 $K_r = \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\| \leq r\}$. The continuity of f guarantees a global minimum $\boldsymbol{x}^\star \in K_r$ with $f(\boldsymbol{x}^\star) \leq f(\mathbf{0})$. It is then $f(\boldsymbol{x}^\star) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^d$ trivially. □

3. Characterization

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

THEOREM 2.14 (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 of a properly differentiable function, we take into account the behavior of the function around $f(\boldsymbol{x})$ with respect to the tangent hyperplane at the point $\big(\boldsymbol{x}, f(\boldsymbol{x})\big)$. Second derivatives make this process very easy.

THEOREM 2.15. *Suppose $f\colon \mathbb{R}^d \to \mathbb{R}$ is coercive and continuously differentiable at a point $\boldsymbol{x}^\star$. If $\boldsymbol{x}^\star$ is a global minimum, then $\nabla f(\boldsymbol{x}^\star) = \mathbf{0}$.*

THEOREM 2.16 (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.

- *If $\boldsymbol{x}^\star$ is a strict local minimum, then $\nabla f(\boldsymbol{x}^\star) = 0$ and $\operatorname{Hess} f(\boldsymbol{x}^\star)$ is* positive definite.

THEOREM 2.17 (Second order sufficient optimality conditions for minimization). *Suppose $f\colon D \subseteq \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable at a point $\boldsymbol{x}^\star$ in the interior of D and $\nabla f(\boldsymbol{x}^\star) = 0$. Then $\boldsymbol{x}^\star$ is a:*

Local Minimum: *if* $\operatorname{Hess} f(\boldsymbol{x}^\star)$ *is* positive semidefinite.
Strict Local Minimum: *if* $\operatorname{Hess} f(\boldsymbol{x}^\star)$ *is* positive definite.
If $D = \mathbb{R}^d$ and $\boldsymbol{x}^\star \in \mathbb{R}^d$ satisfies $\nabla f(\boldsymbol{x}^\star) = 0$, then $\boldsymbol{x}^\star$ is a:
Global Minimum: *if* $\operatorname{Hess} f(\boldsymbol{x})$ *is* positive semidefinite *for all $\boldsymbol{x} \in \mathbb{R}^d$.*
Strict Global Minimum: *if* $\operatorname{Hess} f(\boldsymbol{x})$ *is* positive definite *for all $\boldsymbol{x} \in \mathbb{R}^d$.*

THEOREM 2.18. *Any local minimum of a convex function $f\colon C \to \mathbb{R}$ on a convex set $C \subseteq \mathbb{R}^d$ is also a global minimum. If f is a strictly convex function, then any local minimum is the unique strict global minimum.*

THEOREM 2.19. *Suppose $f\colon C \to \mathbb{R}$ is a convex function with continuous first partial derivatives on a convex set $C \subseteq \mathbb{R}^d$. Then, any critical point of f in C is a global minimum of f.*

Examples

EXAMPLE 2.10. Find a global minimum in $\mathbb{R}^3$ (if it exists) for the function

$$f(x, y, z) = e^{x-y} + e^{y-x} + e^{x^2} + z^2.$$

This function has continuous partial derivatives of any order in $\mathbb{R}^3$. Its continuity does not guarantee existence of a global minimum initially since the domain is not compact, but we may try our luck with its critical points. Note $\nabla f(x, y, z) = \left[e^{x-y} - e^{y-x} + 2xe^{x^2}, -e^{x-y} + e^{y-x}, 2x\right]$. The only critical point is then $(0, 0, 0)$ (Why?). The Hessian at that point is positive definite:

$$\operatorname{Hess} f(0, 0, 0) = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \Delta_1 = 4 > 0, \quad \Delta_2 = 4 > 0, \quad \Delta_3 = 8 > 0.$$

By Theorem 2.17, $f(0, 0, 0) = 3$ is a priori a strict local global minimum value. To prove that this point is actually a strict global minimum, notice that

$$\operatorname{Hess} f(x, y, z) = \begin{bmatrix} e^{x-y} + e^{y-x} + 4x^2e^{x^2} + 2e^{x^2} & -e^{x-y} - e^{y-x} & 0 \\ -e^{x-y} - e^{y-x} & e^{x-y} + e^{y-x} & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The first principal minor is trivially positive: $\Delta_1 = e^{x-y} + e^{y-x} + 4x^2e^{x^2} + 2e^{x^2}$, since it is a sum of three positive terms and on non-negative term. The second principal minor is also positive:

$$\Delta_2 = \det \begin{bmatrix} e^{x-y} + e^{y-x} + 4x^2e^{x^2} + 2e^{x^2} & -e^{x-y} - e^{y-x} \\ -e^{x-y} - e^{y-x} & e^{x-y} + e^{y-x} \end{bmatrix}$$

$$= (e^{x-y} + e^{y-x})^2 + (e^{x-y} + e^{y-x})(4x^2e^{x^2} + 2e^{x^2}) - (e^{x-y} + e^{y-x})^2$$
$$= (e^{x-y} + e^{y-x})(4x^2e^{x^2} + 2e^{x^2}) > 0$$

The third principal minor is positive too: $\Delta_3 = 2\Delta_2 > 0$. We have just proved that $\mathrm{Hess} f(x, y, z)$ is positive definite for all $(x, y, z) \in \mathbb{R}^3$, and thus $(0, 0, 0)$ is a strict global minimum.

EXAMPLE 2.11. Find global minima in $\mathbb{R}^2$ (if they exist) for the function

$$f(x, y) = e^{x-y} + e^{y-x}.$$

This function also has continuous partial derivatives of any order, but no extrema is guaranteed a priori. Notice that all points (x, y) satisfying $y = x$ are critical. For such points, the corresponding Hessians and principal minors are given by

$$\mathrm{Hess} f(x, x) = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}, \qquad \Delta_1 = 2 > 0, \qquad \Delta_2 = 0;$$

therefore, $\mathrm{Hess} f(x, x)$ is positive semidefinite for each critical point. By Theorem 2.17, $f(x, x) = 2$ is a local minimum for all $x \in \mathbb{R}$. To prove they are global minima, notice that for each $(x, y) \in \mathbb{R}^2$:

$$\mathrm{Hess} f(x, y) = \begin{bmatrix} e^{x-y} + e^{y-x} & -e^{x-y} - e^{y-x} \\ -e^{x-y} - e^{y-x} & e^{x-y} + e^{y-x} \end{bmatrix},$$
$$\Delta_1 = e^{x-y} + e^{y-x} > 0, \quad \Delta_2 = 0.$$

The Hessian is positive semidefinite for all points, hence proving that any point in the line $y = x$ is a global minimum of f.

EXAMPLE 2.12. Find local and global minima in $\mathbb{R}^2$ (if they exist) for the function

$$f(x, y) = x^3 - 12xy + 8y^3.$$

This is a polynomial of degree 3, so we have continuous partial derivatives of any order. It is easy to see that this function has no global minima:

$$\lim_{x \to -\infty} f(x, 0) = \lim_{x \to -\infty} x^3 = -\infty.$$

Let's search instead for local minima. From the equation $\nabla f(x, y) = \mathbf{0}$ we obtain two critical points: $(0, 0)$ and $(2, 1)$. The corresponding Hessians and their eigenvalues are:

$$\mathrm{Hess} f(0, 0) = \begin{bmatrix} 0 & -12 \\ -12 & 0 \end{bmatrix}, \quad \lambda_1 = -12 < 0, \quad \lambda_2 = 12 > 0,$$
$$\mathrm{Hess} f(2, 1) = \begin{bmatrix} 12 & -12 \\ -12 & 48 \end{bmatrix}, \quad \lambda_1 = 30 - 6\sqrt{13} > 0, \quad \lambda_2 = 30 + 6\sqrt{30} > 0.$$

By Theorem 2.17, we have that $f(2, 1) = -8$ is a local minimum, but $f(0, 0) = 0$ is not.

EXAMPLE 2.13. Find local and global minima in $\mathbb{R}^2$ (if they exist) for the function

$$f(x,y) = x^4 - 4xy + y^4.$$

This is a polynomial of degree 4, so we do have continuous partial derivatives of any order. There are three critical points: $(0,0)$, $(-1,-1)$ and $(1,1)$. The latter two are both strict local minima (by virtue of Theorem 2.17).

$$\text{Hess}f(-1,-1) = \text{Hess}f(1,1) = \begin{bmatrix} 12 & -4 \\ -4 & 12 \end{bmatrix}, \quad \Delta_1 = 12 > 0, \quad \Delta_2 = 128 > 0.$$

We proved in Example 2.6 that f is coercive. By Theorems 2.13 and 2.15 we have that $f(-1,-1) = f(1,1) = -2$ must be strict global minimum values.

EXAMPLE 2.14. Find local and global minima in $\mathbb{R}^2$ (if they exist) for the function

$$f(x,y) = 2x^2 + y^2 + \frac{1}{2x^2 + y^2}.$$

The simplest way is to realize that f is a convex function on $\mathbb{R}^2$. We prove this by re-writing $f = g \circ h$ as a composition of $g(x) = x + 1/x$ on $(0,\infty)$ with $h(x,y) = 2x^2 + y^2$ on $\mathbb{R}^2$. Note that both functions are strictly convex (Why?).

The only critical point of g in $(0,\infty)$ lies at $x = 1$. By Theorem 2.19, $g(1) = 2$ must be a strict global minimum value of g. That being the case, any point (x,y) satisfying $h(x,y) = 1$ is a global minimum of f. These are all the points on the ellipse $2x^2 + y^2 = 1$.

Exercises

PROBLEM 2.1 (Basic). Consider the function

$$f(x,y) = \frac{x+y}{2+\cos x}$$

At what points $(x,y) \in \mathbb{R}^2$ is this function continuous?

PROBLEM 2.2 (Intermediate). Give an example of a 2×2 symmetric matrix of each kind below:

(a) positive definite,
(b) positive semidefinite,
(c) negative definite,
(d) negative semidefinite,
(e) indefinite.

PROBLEM 2.3 (Basic). [**7**, p.31, #2] Classify the following matrices according to whether they are positive or negative definite or semidefinite or indefinite:

$$\text{(a)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -2 \end{bmatrix} \qquad \text{(c)} \begin{bmatrix} 7 & 0 & 0 \\ 0 & -8 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

(d) $\begin{bmatrix} 3 & 1 & 2 \\ 1 & 5 & 3 \\ 2 & 3 & 7 \end{bmatrix}$ (e) $\begin{bmatrix} -4 & 0 & 1 \\ 0 & -3 & 2 \\ 1 & 2 & -5 \end{bmatrix}$ (f) $\begin{bmatrix} 2 & -4 & 0 \\ -4 & 8 & 0 \\ 0 & 0 & -3 \end{bmatrix}$

PROBLEM 2.4 (Basic). [**7**, p.31, #3] Write the quadratic form $\mathcal{Q}_{\boldsymbol{A}}(\boldsymbol{x})$ associated with each of the following matrices $\boldsymbol{A}$:

(a) $\begin{bmatrix} -1 & 2 \\ 2 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & -2 & 2 \\ 0 & 2 & 3 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & -3 \\ -3 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} -3 & 1 & 2 \\ 1 & 2 & -1 \\ 2 & -1 & 4 \end{bmatrix}$

PROBLEM 2.5 (Basic). [**7**, p.32, #4] Write each of the quadratic forms below in the form $\boldsymbol{x}\boldsymbol{A}\boldsymbol{x}^{\mathsf{T}}$ for an appropriate symmetric matrix $\boldsymbol{A}$:

(a) $3x^2 - xy + 2y^2$.
(b) $x^2 + 2y^2 - 3z^2 + 2xy - 4xz + 6yz$.
(c) $2x^2 - 4z^2 + xy - yz$.

PROBLEM 2.6 (Intermediate). 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) $f(x, y) = x^4 - 3xy + y^4$.
(d) Rosenbrock functions $\mathcal{R}_{a,b}$.

PROBLEM 2.7 (Advanced). [**7**, p.36, #32] Find an example of a continuous, real-valued, non-coercive function $f\colon \mathbb{R}^2 \to \mathbb{R}$ that satisfies, for all $t \in \mathbb{R}$,

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

PROBLEM 2.8 (Basic). [**7**, p.77, #1,2,7] Determine whether the given functions are convex, concave, strictly convex or strictly concave on the specified domains:

(a) $f(x) = \log(x)$ on $(0, \infty)$.
(b) $f(x) = e^{-x}$ on $\mathbb{R}$.
(c) $f(x) = |x|$ on $[-1, 1]$.
(d) $f(x) = |x^3|$ on $\mathbb{R}$.
(e) $f(x, y) = 5x^2 + 2xy + y^2 - x + 2x + 3$ on $\mathbb{R}^2$.
(f) $f(x, y) = x^2/2 + 3y^2/2 + \sqrt{3}xy$ on $\mathbb{R}^2$.
(g) $f(x, y) = 4e^{3x-y} + 5e^{x^2+y^2}$ on $\mathbb{R}^2$.
(h) $f(x, y, z) = x^{1/2} + y^{1/3} + z^{1/5}$ on $C = \{(x, y, z) : x > 0, y > 0, z > 0\}$.

PROBLEM 2.9 (Intermediate). [**7**, p.79 #11] Sketch the epigraph of the following functions

(a) $f(x) = e^x$.
(b) $f(x, y) = x^2 + y^2$.

PROBLEM 2.10 (Advanced). Prove the Bounded Value and Extreme Value Theorems (Theorems 2.11 and 2.12).

PROBLEM 2.11 (Intermediate). For the following optimization problems, state whether existence of a solution is guaranteed:

(a) $f(x) = (1 + x)/x$ over $[1, \infty)$
(b) $f(x) = 1/x$ over $[1, 2)$
(c) The following piecewise function over $[1, 2]$

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

PROBLEM 2.12 (Advanced). State and prove equivalent results to Theorems 2.14, 2.16 and 2.17 to describe necessary and sufficient conditions for the characterization of *maxima*.

PROBLEM 2.13 (Basic). [**7**, p.32, #7] Use the *Principal Minor Criteria* (Theorem 2.3) to determine—if possible—the nature of the critical points of the following functions:

(a) $f(x, y) = x^3 + y^3 - 3x - 12y + 20$.
(b) $f(x, y, z) = 3x^2 + 2y^2 + 2z^2 + 2xy + 2yz + 2xz$.
(c) $f(x, y, z) = x^2 + y^2 + z^2 - 4xy$.
(d) $f(x, y) = x^4 + y^4 - x^2 - y^2 + 1$.
(e) $f(x, y) = 12x^3 + 36xy - 2y^3 + 9y^2 - 72x + 60y + 5$.

PROBLEM 2.14 (Intermediate). [**7**, p.35 #26] Show that the function

$$f(x, y, z) = e^{x^2+y^2+z^2} - x^4 - y^6 - z^6$$

has a global minimum on $\mathbb{R}^3$.

PROBLEM 2.15 (Intermediate). [**7**, p.36 #33] Consider the function

$$f(x, y) = x^3 + e^{3y} - 3xe^y.$$

Show that f has exactly one critical point, and that this point is a local minimum but not a global minimum.

PROBLEM 2.16 (Basic). Let $f(x, y) = -\log(1 - x - y) - \log x - \log y$.

(a) Find the domain D of f.
(b) Prove that D is a convex set.
(c) Prove that f is strictly convex on D.
(d) Find the strict global minimum.

PROBLEM 2.17 (Basic). [**7**, p.81 #27] Find local and global minima in $\mathbb{R}^3$ (if they exist) for the function

$$f(x, y) = e^{x+z-y} + e^{y-x-z}$$

CHAPTER 3

Numerical Approximation for Unconstrained Optimization

Although technically any characterization result finds the exact value of the extrema of a function, computationally this is hardly feasible (specially for functions of very high dimension). See the following session based on problem 2.14 for an example, where we try to find the critical points of the function $f(x,y,z) = e^{x^2+y^2+z^2} - x^4 - y^y - z^6$ symbolically in Python with the `sympy` libraries:

```
# Importing necessary symbols/libraries/functions
from sympy.abc import x,y,z
from sympy import Matrix, solve, exp
from sympy.tensor.array import derive_by_array

# Description of f, computation of its gradient and Hessian
f = exp(x**2 + y**2 + z**2) - x**4 -y**6 - z**6
gradient = derive_by_array(f, [x,y,z])
hessian  = Matrix([derive_by_array(gradient, a) for a in [x,y,z]])
```

While the correct expressions for ∇f and $\mathrm{Hess} f$ are quickly computed, trying to find critical points results in an error:

```
>>> solve(gradient) # Search of critical points by solving ∇f = 0
NotImplementedError: could not solve
4*x**2*sqrt(-log(exp(x**2)/(2*x**2))) - 6*(-log(exp(x**2)/(2*x**2)))**(5/2)
```

Too complex a task to be performed symbolically, although the obvious answer is $(0,0,0)$. A better way to approach this is by trying to approximate this minimum using the structure of the graph of f. In these notes we are going to explore several strategies to accomplish this task, based on the concept of *iterative methods for finding zeros of real-valued functions*.

1. Newton-Raphson's Method

1.1. Newton-Raphson Method to search for roots of univariate functions. In order to find a good estimation of $\sqrt{2}$ with many decimal places, we allow a computer to find better and better approximations of the root of the polynomial $p(x) = x^2 - 2$. We start with an initial guess, say $x_0 = 3$. We construct a sequence $\{x_n\}_{n\in\mathbb{N}}$ that converges to $\sqrt{2}$ as follows:

(a) Find the tangent line to the graph of p at x_0,

$$y - p(x_0) = p'(x_0)(x - x_0)$$

(b) Provided this line is not horizontal ($p'(x_0) \neq 0$), report the intersection of this line with the x–axis. Call this intersection x_1

$$x_1 = x_0 - \frac{p(x_0)}{p'(x_0)}$$

(c) Repeat this process, to get the sequence

$$x_{n+1} = x_n - \frac{p(x_n)}{p'(x_n)} = x_n - \frac{x_n^2 - 2}{2x_n} = \frac{x_n}{2} - \frac{1}{x_n}.$$

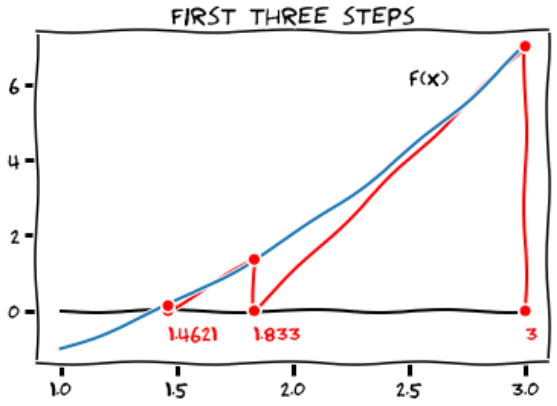

FIGURE 3.1. Newton-Raphson iterative method

Note the result of applying this process a few times:

n	x_n	$p(x_n)$
0	3.000000000000000	$7.0000E+00$
1	1.833333333333333	$1.3611E+00$
2	1.462121212121212	$1.3780E-01$
3	1.414998429894803	$2.2206E-03$
4	1.414213780047198	$6.1568E-07$
5	1.414213562373112	$4.7518E-14$
6	1.414213562373095	$-4.4409E-16$
7	1.414213562373095	$4.4409E-16$

DEFINITION. Given a differentiable real-valued function $f\colon \mathbb{R} \to \mathbb{R}$ and an initial guess $x_0 \in \mathbb{R}$, we define the *Newton-Raphson iteration* to be the sequence given by the following *recursive formula*

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

The *Newton-Raphson method* refers to employing this sequence to search and approximate roots of the equation $f(x) = 0$.

EXAMPLE 3.1. Consider now the function $f(x) = 1 - \frac{1}{x}$ over $(0, \infty)$, which has the obvious root $x = 1$. The Newton-Raphson method gives the following iterates for any $x_0 \in (0, \infty)$:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n\big(2 - x_n\big).$$

Notice the two factors in the right-hand side of that expression: x_n, and $2 - x_n$. If the initial guess does not satisfy $0 < x_0 < 2$, then the next iteration gives a non-positive value (see Figure 3.2). The method will not work on those instances: convergence to a solution is not guaranteed.

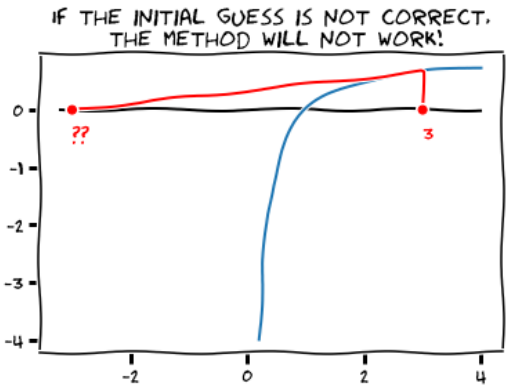

FIGURE 3.2. Initial guess must carefully be chosen in Newton-Raphson

EXAMPLE 3.2. Consider now $f(x) = \operatorname{sign}(x)\sqrt{|x|}$ over $\mathbb{R}$, with root at $x = 0$. The Newton-Raphson method fails miserably with this function: for any $x_0 \neq 0$

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = x_0 - \frac{\operatorname{sign}(x_0)|x_0|^{1/2}}{\frac{1}{2}|x_0|^{-1/2}} = -x_0.$$

This sequence turns into a loop: $x_{2n} = x_0$, $x_{2n+1} = -x_0$ for all $n \in \mathbb{N}$ (see Figure 3.3).

1.2. Efficiency of Newton-Raphson's Method. To study the error in a Newton-Raphson iteration $\{x_n\}_{n\in\mathbb{N}}$ that converges to a root $x^\star$ of the function f, we observe that

$$\begin{aligned} x_{n+1} - x^\star &= x_n - x^\star - \frac{f(x_n)}{f'(x_n)} \\ &= (x_n - x^\star)\left(1 - \frac{f(x_n) - \overbrace{f(x^\star)}^{0}}{(x_n - x^\star)f'(x_n)}\right) \end{aligned}$$

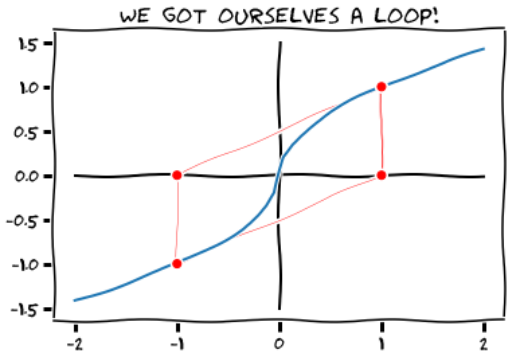

FIGURE 3.3. Newton-Raphson fails for some functions

$$= (x_n - x^\star)\left(1 - \frac{1}{f'(x_n)}\frac{f(x_n) - f(x^\star)}{x_n - x^\star}\right) \tag{3}$$

Recall at this point the definition of *divided differences* for any n–times differentiable function $g\colon \mathbb{R} \to \mathbb{R}$, and a family of values $t_0 \leq t_1 \leq \cdots \leq t_n$:

$$\begin{aligned}
\Delta g[t_0, t_1] &= \frac{g(t_1) - g(t_0)}{t_1 - t_0} \qquad (\text{if } t_0 \neq t_1)\\
\Delta g[t_0, t_0] &= g'(t_0)\\
\Delta g[t_0, t_1, \ldots, t_n] &= \frac{\Delta g[t_1, t_2, \ldots, t_n] - \Delta g[t_0, t_1, \ldots, t_{n-1}]}{t_n - t_0} \quad (\text{if } t_0 \neq t_n)\\
\Delta g[\underbrace{t_0, \ldots, t_0}_{n+1 \text{ times}}] &= \frac{1}{n!} g^{(n)}(t_0) \qquad (\text{Why? } \textbf{Hint:} \text{ Taylor's Polynomial})
\end{aligned}$$

We can then rewrite (3) in terms of divided differences as follows:

$$\begin{aligned}
x_{n+1} - x^\star &= (x_n - x^\star)\left(1 - \frac{\Delta f[x_n, x^\star]}{\Delta f[x_n, x_n]}\right)\\
&= (x_n - x^\star)\frac{\Delta f[x_n, x_n] - \Delta f[x_n, x^\star]}{\Delta f[x_n, x_n]}\\
&= (x_n - x^\star)\frac{\Delta f[x_n, x_n] - \Delta f[x_n, x^\star]}{x_n - x^\star}\frac{x_n - x^\star}{\Delta f[x_n, x_n]}\\
&= (x_n - x^\star)^2\frac{\Delta f[x_n, x_n, x^\star]}{\Delta f[x_n, x_n]}
\end{aligned}$$

Therefore,

$$\lim_n \frac{x_{n+1} - x^\star}{(x_n - x^\star)^2} = \lim_n \frac{\Delta f[x_n, x_n, x^\star]}{\Delta f[x_n, x_n]} = \frac{f''(x^\star)}{2f'(x^\star)}.$$

If $f''(x^\star) \neq 0$, the Newton-Raphson's iteration exhibits quadratic convergence.[1]

[1] See Appendix B

REMARK 3.1. We have just proven that, if a Newton-Raphson iteration for a function f gives a convergent sequence, the convergence is quadratic. But, how can we guarantee convergence to a root of f? The key is in *how far can we start the sequence* given the structure of the graph of f.

THEOREM 3.1 (Local Convergence for the Newton-Raphson Method). *Let $x^\star$ be a simple root of the equation $f(x) = 0$, and there exists $\varepsilon > 0$ so that*

- *f is twice continuously differentiable in the interval $(x^\star - \varepsilon, x^\star + \varepsilon)$, and*
- *there are no critical points of f on that interval.*

Set

$$M(\varepsilon) = \max\left\{\left|\frac{f''(s)}{2f'(t)}\right| : x^\star - \varepsilon < s, t < x^\star + \varepsilon\right\}.$$

If ε is small enough so that $\varepsilon M(\varepsilon) < 1$, then

(a) There are no other roots of f in $(x^\star - \varepsilon, x^\star + \varepsilon)$.
(b) Any Newton-Raphson iteration starting at an initial guess $x_0 \neq x^\star$ in that interval will converge (quadratically) to $x^\star$

I have seen this Theorem in [**5**] *with the condition $2\varepsilon M(\varepsilon) < 1$ instead, but I could not see why that 2 was necessary. What am I missing?*

PROOF. Start with Taylor's Theorem for f around $x^\star$. Given $x \neq x^\star$ satisfying $|x - x^\star| < \varepsilon$, there exists ξ between x and $x^\star$ so that

$$\begin{aligned} f(x) &= f(x^\star) + (x - x^\star)f'(x^\star) + \tfrac{1}{2}(x - x^\star)^2 f''(\xi) \\ &= (x - x^\star)f'(x^\star)\left(1 + (x - x^\star)\frac{f''(\xi)}{2f'(x^\star)}\right) \end{aligned}$$

Note that the three factors on the last expression are never zero:

$$\begin{aligned} x - x^\star &\neq 0 && \text{(since } x \neq x^\star \text{ by hypothesis)} \\ f'(x^\star) &\neq 0 && \text{(no critical points by hypothesis)} \\ \left|(x - x^\star)\frac{f''(\xi)}{2f'(x^\star)}\right| &\leq \varepsilon M(\varepsilon) < 1 && \text{(by hypothesis on } M(\varepsilon)) \end{aligned}$$

This proves (a).

We want to prove now that all terms of a Newton-Raphson iteration stay in the interval $(x^\star - \varepsilon, x^\star + \varepsilon)$. We do that by induction:

- $|x_0 - x^\star| < \varepsilon$ by hypothesis.
- Assume $|x_n - x^\star| < \varepsilon$. In that case,

$$\left|x_{n+1} - x^\star\right| = \left|x_n - x^\star\right|^2 \left|\frac{\Delta f[x_n, x_n, x^\star]}{\Delta f[x_n, x_n]}\right|$$

but $\Delta f[x_n, x_n] = f'(x_n)$, and there exist ξ_n between x_n and $x^\star$ so that $\Delta f[x_n, x_n, x^\star] = \frac{1}{2}f''(\xi_n)$; therefore,

$$\left|x_{n+1} - x^\star\right| = \left|x_n - x^\star\right|^2 \left|\frac{f''(\xi)}{2f'(x_n)}\right| \leq \varepsilon^2 M(\varepsilon) = \varepsilon \cdot \varepsilon M(\varepsilon) < \varepsilon.$$

The next step is to prove that there is convergence. A similar computation to the previous gives

$$\big|x_n - x^\star\big| \leq \varepsilon M(\varepsilon)\big|x_{n-1} - x^\star\big| \leq \big(\varepsilon M(\varepsilon)\big)^n\big|x_0 - x^\star\big|.$$

Since $\varepsilon M(\varepsilon) < 1$, $\lim_n \big(\varepsilon M(\varepsilon)\big)^n = 0$, and $\{x_n\}_{n\in\mathbb{N}}$ converges to $x^\star$. □

1.3. Extension to higher dimensions. Let's proceed to extend this process to functions $\boldsymbol{g}\colon \mathbb{R}^d \to \mathbb{R}^d$ as follows.

- Any function $\boldsymbol{g}\colon \mathbb{R}^d \to \mathbb{R}^d$ can be described in the form $\boldsymbol{g}(\boldsymbol{x}) = \big[g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \ldots, g_d(\boldsymbol{x})\big]$ for d real-valued functions $g_k\colon \mathbb{R}^d \to \mathbb{R}$ $(1 \leq k \leq d)$.
- For such a function $\boldsymbol{g}$, we may express its gradient as a $d \times d$ matrix in the form

$$\nabla \boldsymbol{g} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_d} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_d}{\partial x_1} & \frac{\partial g_d}{\partial x_2} & \cdots & \frac{\partial g_d}{\partial x_d} \end{bmatrix}$$

Start with a guess for the solution, $\boldsymbol{x}_0$, and on the n–th step of the algorithm compute the $(n+1)$–th term of the sequence by

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \big[\nabla \boldsymbol{g}(\boldsymbol{x}_n)\big]^{-1}\boldsymbol{g}(\boldsymbol{x}_n), \tag{4}$$

where $\big[\nabla \boldsymbol{g}(\boldsymbol{x}_n)\big]^{-1}$ represents the inverse matrix of the gradient at $\boldsymbol{x}_n$. This is equivalent to selecting in the tangent hyperplane to the graph of $\boldsymbol{g}$ at $\boldsymbol{g}(\boldsymbol{x}_n)$, the one line in the direction with the most rapid increase/decrease. The computation of $\boldsymbol{x}_{n+1}$ is therefore the intersection of that line with the hyperplane $x_d = 0$. We refer to $\big[\nabla \boldsymbol{g}(\boldsymbol{x}_n)\big]^{-1}\boldsymbol{g}(\boldsymbol{x}_n)$ as the *Newton-Raphson direction* for $\boldsymbol{g}$ at $\boldsymbol{x}_n$.

EXAMPLE 3.3. Consider the function $\boldsymbol{g}\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\boldsymbol{g}(x, y, z) = \big[x^3 - y, y^3 - x\big]$$

Its gradient at each (x, y) is given by

$$\nabla \boldsymbol{g}(x, y) = \begin{bmatrix} 3x^2 & -1 \\ -1 & 3y^2 \end{bmatrix}$$

Note the determinant of this matrix is $\det \nabla \boldsymbol{g}(x, y) = 9x^2y^2 - 1 = (3xy - 1)(3xy + 1)$. For any point (x, y) that does not make this expression zero, this is an invertible matrix with

$$\big[\nabla \boldsymbol{g}(x, y)\big]^{-1} = \frac{1}{9x^2y^2 - 1}\begin{bmatrix} 3y^2 & 1 \\ 1 & 3x^2 \end{bmatrix}$$

For an initial guess (x_0, y_0), the sequence computed by this method is then given by

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix} - \frac{1}{9x_n^2 y_n^2 - 1} \begin{bmatrix} 3y_n^2 & 1 \\ 1 & 3x_n^2 \end{bmatrix} \begin{bmatrix} x_n^3 - y_n \\ y_n^3 - x_n \end{bmatrix}$$

Let's run this process with three different initial guesses:

(a) Starting at $(x_0, y_0) = (-1.0, 1.0)$, the sequence converges to $(0, 0)$.

n	x_n	y_n
0	−1.00000000	1.00000000
1	−0.50000000	0.50000000
2	−0.14285714	0.14285714
3	−0.00549451	0.00549451
4	−0.00000033	0.00000033
5	−0.00000000	0.00000000
6	−0.00000000	0.00000000

(b) Starting at $(x_0, y_0) = (3.5, 2.1)$, the sequence converges to $(1, 1)$.

n	x_n	y_n
0	3.50000000	2.10000000
1	2.37631607	1.57961573
2	1.65945969	1.27476534
3	1.23996276	1.10419072
4	1.04837462	1.02274752
5	1.00260153	1.00133122
6	1.00000824	1.00000451
7	1.00000000	1.00000000
8	1.00000000	1.00000000

(c) Starting at $(x_0, y_0) = (-13.5, -7.3)$, the sequence converges to $(-1, -1)$.

n	x_n	y_n	n	x_n	y_n
0	−13.50000000	−7.30000000	7	−1.09518303	−1.04341362
1	−9.00900415	−4.92301873	8	−1.00932090	−1.00463507
2	−6.01982204	−3.36480659	9	−1.00010404	−1.00005571
3	−4.03494126	−2.36199873	10	−1.00000001	−1.00000001
4	−2.72553474	−1.73750959	11	−1.00000000	−1.00000000
5	−1.87830623	−1.36573112	12	−1.00000000	−1.00000000
6	−1.36121191	−1.15374930	13	−1.00000000	−1.00000000

1.4. Optimization via Newton's Method. We can readily see how this process aids in the computation of critical points of twice continuously differentiable real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$:

(a) Set $\boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_d}\right]$

(b) It is then $\nabla \boldsymbol{g}(\boldsymbol{x}) = \operatorname{Hess} f(\boldsymbol{x})$
(c) Perform a Newton method (with initial guess $\boldsymbol{x}_0$) on $\boldsymbol{g} = \nabla f$ to obtain the recurrence formula

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \big[\operatorname{Hess} f(\boldsymbol{x}_n)\big]^{-1} \cdot \nabla f(\boldsymbol{x}_n) \tag{5}$$

EXAMPLE 3.4. Consider the polynomial $p_4(x, y) = x^4 - 4xy + y^4$. Notice $\nabla p_4(x, y) = \big[x^3 - y, y^3 - x\big]$—this is function $\boldsymbol{g}$ in Example 3.3. The critical points we found were $(0, 0)$, $(-1, -1)$ and $(1, 1)$. See Figure 3.4.

EXAMPLE 3.5. A similar process for the Rosenbrock function

$$\mathcal{R}_{1,1}(x, y) = (1 - x)^2 + (y - x^2)^2$$

gives the following recurrence formula:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix} - \big[\operatorname{Hess} \mathcal{R}_{1,1}(x_n, y_n)\big]^{-1} \cdot \nabla \mathcal{R}_{1,1}(x_n, y_n)$$
$$= \frac{1}{2x_n^2 - 2y_n + 1} \begin{bmatrix} 2x_n^3 - 2x_n y_n + 1 \\ x_n(2x_n^3 - 2x_n y_n - x_n + 2) \end{bmatrix}$$

For instance, starting with the initial guess $(x_0, y_0) = (-2, 2)$, the sequence converges to the critical point $(1, 1)$. See Figure 3.4.

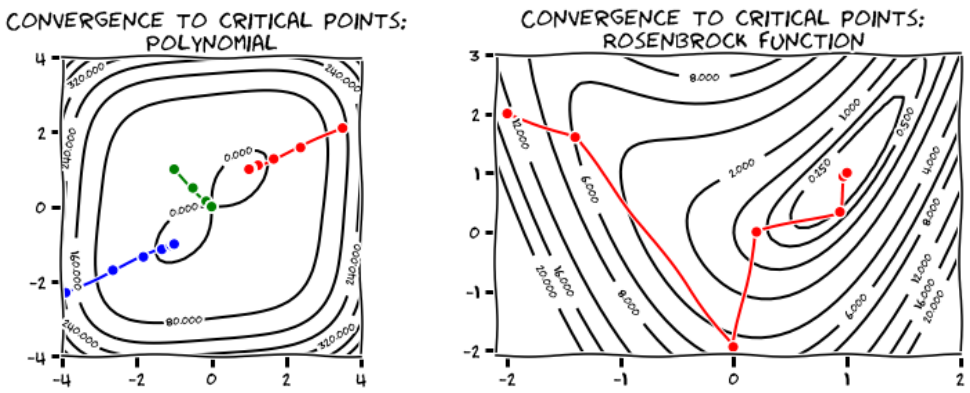

FIGURE 3.4. Newton-Raphson method

REMARK 3.2. The Newton-Raphson's method to solve $\boldsymbol{g} = \boldsymbol{0}$, as given by the recurrence formula in equation (4) in page 30, is very convenient to provide explicit descriptions of the different iterations. However, it is hardly suitable for practical purposes, due to the computational issues involving matrix inversion.

To avoid dealing with matrix inversion, we consider the following equivalent formula:

$$\nabla \boldsymbol{g}(\boldsymbol{x}_n) \cdot \big(\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\big) = -\boldsymbol{g}(\boldsymbol{x}_n) \tag{6}$$

This is a simple system of linear equations, a much more reliable process, prone to less numerical error.

The equivalent recurrence formula to search for critical points of a twice continuously differentiable real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$ is thus

$$\mathrm{Hess} f(\boldsymbol{x}_n) \cdot (\boldsymbol{x}_{n+1} - \boldsymbol{x}_n) = -\nabla f(\boldsymbol{x}_n). \tag{7}$$

EXAMPLE 3.6. The equivalent recurrence formula to the one we obtained in example 3.3 is as follows:

$$\begin{bmatrix} 3x_n^2 & -1 \\ -1 & 3y_n^2 \end{bmatrix} \begin{bmatrix} x_{n+1} - x_n \\ y_{n+1} - y_n \end{bmatrix} = \begin{bmatrix} x_n^3 - y_n \\ y_n^3 - x_n \end{bmatrix}$$

All we need to do is, at each step n, solve for X and Y the system of linear equations

$$\begin{cases} 3x_n^2(X - x_n) - (Y - y_n) = x_n^3 - y_n \\ -(X - x_n) + 3y_n^2(Y - y_n) = y_n^2 - x_n \end{cases}$$

or equivalently,

$$\begin{cases} 3x_n^2 X - Y = 4x_n^3 - 2y_n \\ -X + 3y_n^2 Y = 4y_n^2 - 2x_n \end{cases}$$

There are some theoretical results that aid in the search for a *good* initial guess in case of multivariate functions. The following states a simple set of conditions on f and $\boldsymbol{x}_0$ to guarantee *quadratic convergence* of the corresponding sequences $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ to a critical point $\boldsymbol{x}^\star$.

THEOREM 3.2 (Quadratic Convergence Theorem). *Suppose $f\colon \mathbb{R}^d \to \mathbb{R}$ is a twice continuously differentiable real-valued function, and $\boldsymbol{x}^\star$ is a critical point of f. Let $\mathcal{N}(\boldsymbol{x}) = \boldsymbol{x} - \big[\mathrm{Hess} f(\boldsymbol{x})\big]^{-1} \cdot \nabla f(\boldsymbol{x})$. If there exists*

(a) $h > 0$ so that[2] $\big\|\big[\mathrm{Hess} f(\boldsymbol{x}^\star)\big]^{-1}\big\| \le \frac{1}{h}$,
(b) $\beta > 0$, $L > 0$ for which $\|\mathrm{Hess} f(\boldsymbol{x}) - \mathrm{Hess} f(\boldsymbol{x}^\star)\| \le L\|\boldsymbol{x} - \boldsymbol{x}^\star\|$ provided $\|\boldsymbol{x} - \boldsymbol{x}^\star\| \le \beta$.

In that case, for all $\boldsymbol{x} \in \mathbb{R}^d$ satisfying $\|\boldsymbol{x} - \boldsymbol{x}^\star\| \le \min\{\beta, \frac{2h}{3L}\}$,

$$\frac{\|\mathcal{N}(\boldsymbol{x}) - \boldsymbol{x}^\star\|}{\|\boldsymbol{x} - \boldsymbol{x}^\star\|^2} \le \frac{3L}{2h}$$

2. The Method of Steepest Descent

The method of Steepest Descent is based upon the following property of gradients that we learned in Vector Calculus:

THEOREM 3.3. *If $f\colon \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable, then at any point $\boldsymbol{x} \in \mathbb{R}^d$, the vector $-\nabla f(\boldsymbol{x})$ points in the direction of most rapid decrease for f at $\boldsymbol{x}$. The rate of decrease of f at $\boldsymbol{x}$ in this direction is precisely $-\|\nabla f(\boldsymbol{x})\|$.*

[2]Recall the *norm* of a matrix M, defined by $\|M\| = \max\{\|M \cdot \boldsymbol{x}\| : \|\boldsymbol{x}\| = 1\}$.

REMARK 3.3. For this reason, the vector $-\nabla f(\boldsymbol{x})$ is called the *direction of steepest descent* of f at $\boldsymbol{x}$.

In order to search for a local minimum for a twice continuously differentiable function $f\colon \mathbb{R}^d \to \mathbb{R}$, we start by choosing an initial guess $\boldsymbol{x}_0$.

(a) Restrict the function f over the line through $\boldsymbol{x}_0$ in the direction of $-\nabla f(\boldsymbol{x}_0)$:

$$\varphi_0(t) = f\big(\boldsymbol{x}_0 - t\,\nabla f(\boldsymbol{x}_0)\big), \quad t \geq 0$$

(b) Search for the value of $t_0 \geq 0$ that minimizes φ_0, and set

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 - t_0\,\nabla f(\boldsymbol{x}_0)$$

(c) Repeat this process to get the sequence

$$\begin{aligned} &\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - t_n\,\nabla f(\boldsymbol{x}_n), \\ &t_n = \operatorname*{arg\,min}_{t\geq 0} \varphi_n(t) = \operatorname*{arg\,min}_{t\geq 0} f\big(\boldsymbol{x}_n - t\,\nabla f(\boldsymbol{x}_n)\big) \end{aligned} \tag{8}$$

REMARK 3.4. Sequences constructed following the formula in (8) are said to be *sequences of Steepest Descent* for f.

Unlike Newton's method, this algorithm guarantees that these sequences are non-increasing: $f(\boldsymbol{x}_{n+1}) \leq f(\boldsymbol{x}_n)$ for all $n \in \mathbb{N}$. And even better: if there is convergence, their limit must a critical point of f. These results are formalized in Theorems 3.4 and 3.5 below.

Steepest descent sequences have another interesting property: on each step n, the direction of descent $x_{n+1} - x_n$ is perpendicular to the direction of the next step! We state and prove this result in Theorem 3.6.

THEOREM 3.4. *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable real-valued function, and let $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ be a sequence of steepest descent for f. If $\nabla f(\boldsymbol{x}_N) \neq 0$, then $f(\boldsymbol{x}_{N+1}) < f(\boldsymbol{x}_N)$.*

THEOREM 3.5. *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a real-valued function, let $\boldsymbol{x}_0 \in \mathbb{R}^d$ be an initial guess. Assume $S = \{\boldsymbol{x} \in \mathbb{R}^d : f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)\}$ is a compact set and f is continuously differentiable in S. Under these conditions, the limit of any convergent subsequence of the associated sequence of steepest descent $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ is a critical point of f.*

THEOREM 3.6. *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable real-valued function, and $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ a sequence of steepest descent for f. For any $n \in \mathbb{N}$, $\langle \boldsymbol{x}_{n+2} - \boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+1} - \boldsymbol{x}_n \rangle = 0$.*

PROOF. Consider for each $n \in \mathbb{N}$ the function $\varphi_n(t) = f\big(\boldsymbol{x}_n - t\,\nabla f(\boldsymbol{x}_x)\big)$, with a global minimum at $t_n \geq 0$. It must then be

$$0 = \varphi_n'(t_n) = \langle \nabla f(\boldsymbol{x}_n), -\nabla f\left(\boldsymbol{x}_n - t_n\,\nabla f(\boldsymbol{x}_n)\right)\rangle = -\langle \nabla f(\boldsymbol{x}_{n+1}), \nabla f(\boldsymbol{x}_n)\rangle,$$

which proves that the gradient of consecutive terms of the sequence of steepest descent for f are perpendicular. Now, by virtue of the recurrence formula

(8),

$$\begin{aligned}\langle \boldsymbol{x}_{n+2} - \boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+1} - \boldsymbol{x}_n \rangle &= \langle t_{n+1} \nabla f(\boldsymbol{x}_{n+1}), t_n \nabla f(\boldsymbol{x}_n) \rangle \\ &= t_{n+1} t_n \langle \nabla f(\boldsymbol{x}_{n+1}), \nabla f(\boldsymbol{x}_n) \rangle = 0,\end{aligned}$$

which proves the statement. □

EXAMPLE 3.7. For the polynomial function $p_4(x, y) = x^4 - 4xy + y^4$ from Example 3.4, using the same initial guesses as in Example 3.3, we find the following behavior:

- Starting at $(x_0, y_0) = (-1.0, 1.0)$, the sequence also converges to $(0, 0)$, but this time in one step.

n	x_n	y_n	$f(x_n, y_n)$
0	−1.000000	1.000000	6.000000
1	0.000000	0.000000	0.000000
2	nan	nan	nan

- Starting at $(x_0, y_0) = (3.5, 2.1)$, the sequence converges to $(1, 1)$.

n	x_n	y_n	$f(x_n, y_n)$	n	x_n	y_n	$f(x_n, y_n)$
0	3.500000	2.100000	140.110600	8	1.000149	1.000067	−2.000000
1	1.044472	1.753064	3.310777	9	1.000010	1.000048	−2.000000
2	1.141931	1.063276	−1.878163	10	1.000015	1.000007	−2.000000
3	1.008581	1.044435	−1.988879	11	1.000001	1.000005	−2.000000
4	1.013966	1.006319	−1.998931	12	1.000002	1.000001	−2.000000
5	1.000898	1.004472	−1.999891	13	1.000000	1.000001	−2.000000
6	1.001437	1.000651	−1.999989	14	1.000000	1.000000	−2.000000
7	1.000093	1.000461	−1.999999	15	1.000000	1.000000	−2.000000

- Starting at $(x_0, y_0) = (-13.5, -7.3)$, the sequence converges to $(1, 1)$ as well.

n	x_n	y_n	$f(x_n, y_n)$	n	x_n	y_n	$f(x_n, y_n)$
0	−13.500000	−7.300000	35660.686600	8	1.000399	1.000185	−1.999999
1	2.362722	−4.871733	640.498302	9	1.000024	1.000127	−2.000000
2	1.434154	1.194162	−0.586492	10	1.000041	1.000019	−2.000000
3	1.021502	1.130993	−1.896212	11	1.000002	1.000013	−2.000000
4	1.038817	1.017881	−1.991558	12	1.000004	1.000002	−2.000000
5	1.002305	1.012291	−1.999167	13	1.000000	1.000001	−2.000000
6	1.003909	1.001808	−1.999917	14	1.000000	1.000000	−2.000000
7	1.000236	1.001246	−1.999992	15	1.000000	1.000000	−2.000000

EXAMPLE 3.8. Notice what happens when we try to implement the same process on the Rosenbrock function $\mathcal{R}_{1,1}(x, y) = (1-x)^2 + (y-x^2)^2$, with the initial guess $(x_0, y_0) = (-2, 2)$. The sequence does converge to the minimum $(1, 1)$, albeit very slowly.

n	x_n	y_n	$f(x_n, y_n)$	n	x_n	y_n	$f(x_n, y_n)$
0	−2.000000	2.000000	13.000000	17	0.916394	0.789239	0.009544
1	−0.166290	2.309522	6.567163	18	0.911201	0.818326	0.008028
2	0.256054	−0.056128	0.568264	19	0.929317	0.821560	0.006766
3	0.613477	0.007683	0.285318	20	0.925024	0.845608	0.005723
4	0.568566	0.259241	0.190235	21	0.939976	0.848277	0.004847
5	0.715784	0.285524	0.132227	22	0.936397	0.868329	0.004118
6	0.689755	0.431319	0.098227	23	0.948845	0.870551	0.003502
7	0.779264	0.447299	0.074310	24	0.945840	0.887385	0.002986
8	0.761554	0.546496	0.057977	25	0.956276	0.889248	0.002548
9	0.823325	0.557524	0.045696	26	0.953739	0.903457	0.002178
10	0.810322	0.630358	0.036667	27	0.962537	0.905028	0.001864
11	0.855862	0.638488	0.029614	28	0.960386	0.917075	0.001597
12	0.845883	0.694385	0.024199	29	0.967837	0.918405	0.001369
13	0.880846	0.700627	0.019862	30	0.966007	0.928657	0.001176
14	0.872964	0.744776	0.016437	31	0.972342	0.929788	0.001010
15	0.900551	0.749702	0.013647	32	0.970780	0.938539	0.000869
16	0.894200	0.785276	0.011399	33	0.976182	0.939503	0.000748

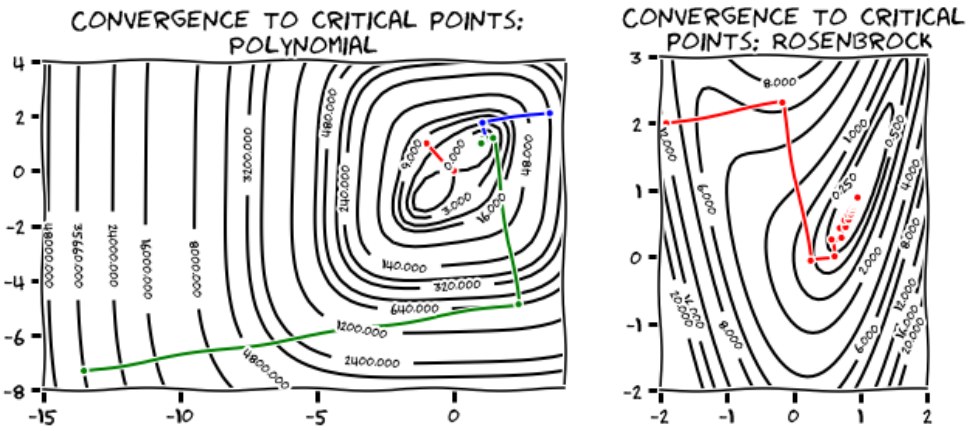

FIGURE 3.5. The Method of Steepest Descent

2.1. Efficiency of Steepest Descent Method. The analysis of efficiency of the method of Steepest descent is quite involved, but it boils down to studying the efficiency of Steepest descent for quadratic functions—since any function can be approximated using a Taylor's polynomial of degree two. We will study that easier case in these notes.

THEOREM 3.7 (Taylor's Formula). *If $f\colon \mathbb{R}^d \to \mathbb{R}$ is a real-valued function of d variables with continuous first and second partial derivatives on $\mathbb{R}^d$, then*

for any choice $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, there exists a point $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{x}, \boldsymbol{y})$ in the segment joining $\boldsymbol{x}$ and $\boldsymbol{y}$ so that

$$f(\boldsymbol{x}) = f(\boldsymbol{y}) + \langle \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle + \tfrac{1}{2}\mathcal{Q}_{\mathbf{Hess}\boldsymbol{f}(\boldsymbol{\xi})}(\boldsymbol{x} - \boldsymbol{y})$$

Assume we have a quadratic function $p\colon \mathbb{R}^d \to \mathbb{R}$ satisfying $p(\mathbf{0}) = 0$. There exist a d-dimensional vector $D = \big[q_1, \ldots, q_d\big]$ and a symmetric matrix $Q = \big[q_{jk}\big]_{j,k=1}^{d}$ (with $q_{jk} = q_{kj}$ for all $1 \leq j, k \leq d$) so that

$$p(\boldsymbol{x}) = \langle D, \boldsymbol{x} \rangle + \tfrac{1}{2}\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{x}) = \sum_{k=1}^{d} \big(\tfrac{1}{2}q_{kk}x_k^2 + q_k x_k\big) + \sum_{1 \leq j < k \leq d} q_{jk} x_j x_k$$

The gradient of this function is thus $\nabla p(\boldsymbol{x}) = x \cdot Q + D$. It has one unique critical point $\boldsymbol{x}^\star = -DQ^{-1}$ (Why?). At that point, it is

$$\begin{aligned} p(\boldsymbol{x}^\star) &= \tfrac{1}{2}\big(-DQ^{-1}\big)Q\big(-DQ^{-1}\big)^\intercal + D\big(-DQ^{-1}\big)^\intercal \\ &= \tfrac{1}{2}DQ^{-1}D^\intercal - DQ^{-1}D^\intercal \\ &= -\tfrac{1}{2}DQ^{-1}D^\intercal = -\tfrac{1}{2}\mathcal{Q}_{(\boldsymbol{Q}^{-1})}(D). \end{aligned} \tag{9}$$

If $\boldsymbol{x}_n$ is a term in a sequence of steepest descent, then to compute $\boldsymbol{x}_{n+1}$ we proceed as follows:

(a) The direction of steepest descent at $\boldsymbol{x}_n$ is

$$\boldsymbol{v}_n = -\nabla p(\boldsymbol{x}_n) = -(x_n Q + D).$$

(b) The restriction $\varphi\colon (0, \infty) \to \mathbb{R}$ of the quadratic function p over the half-line through $\boldsymbol{x}_n$ in the direction $\boldsymbol{v}_n$ is given by

$$\begin{aligned} \varphi(t) &= p(\boldsymbol{x}_n + t\boldsymbol{v}_n) \\ &= \tfrac{1}{2}(\boldsymbol{x}_n + t\boldsymbol{v}_n)Q(\boldsymbol{x}_n + t\boldsymbol{v}_n)^\intercal + D(\boldsymbol{x}_n + t\boldsymbol{v}_n)^\intercal \\ &= \tfrac{1}{2}\boldsymbol{x}_n Q(\boldsymbol{x}_n + t\boldsymbol{v}_n)^\intercal + \tfrac{1}{2}t\boldsymbol{v}_n Q(\boldsymbol{x}_n + t\boldsymbol{v}_n)^\intercal \\ &\quad + D\boldsymbol{x}_n^\intercal + tD\boldsymbol{v}_n^\intercal \\ &= \tfrac{1}{2}\boldsymbol{x}_n Q\boldsymbol{x}_n^\intercal + \tfrac{1}{2}t\boldsymbol{x}_n Q\boldsymbol{v}_n^\intercal + \tfrac{1}{2}t\boldsymbol{v}_n Q\boldsymbol{x}_n^\intercal + \tfrac{1}{2}t^2\boldsymbol{v}_n Q\boldsymbol{v}_n^\intercal \\ &\quad + D\boldsymbol{x}_n^\intercal + tD\boldsymbol{v}_n^\intercal \\ &= \tfrac{1}{2}\underbrace{\boldsymbol{v}_n Q\boldsymbol{v}_n^\intercal}_{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)}t^2 + \underbrace{\tfrac{1}{2}\boldsymbol{x}_n Q\boldsymbol{x}_n^\intercal + D\boldsymbol{x}_n^\intercal}_{p(\boldsymbol{x}_n)} + tD\boldsymbol{v}_n^\intercal + t\boldsymbol{x}_n Q\boldsymbol{v}_n^\intercal \\ &= \tfrac{1}{2}\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)t^2 + p(\boldsymbol{x}_n) + t\underbrace{\big(x_n Q + D\big)}_{-\boldsymbol{v}_n}\boldsymbol{v}_n^\intercal \\ &= \tfrac{1}{2}t^2\boldsymbol{v}_n Q\boldsymbol{v}_n^\intercal - t\boldsymbol{v}_n\boldsymbol{v}_n^\intercal + p(\boldsymbol{x}_n) \\ &= \tfrac{1}{2}\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)t^2 - \|\boldsymbol{v}_n\|^2 t + p(\boldsymbol{x}_n) \end{aligned}$$

(c) The restriction function has its global minimum at

$$t_n = \frac{\|\boldsymbol{v}_n\|^2}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)};$$

therefore, the next iteration occurs at

$$x_{n+1} = x_n + t_n \boldsymbol{v}_n = x_n + \frac{\|\boldsymbol{v}_n\|^2}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)} \boldsymbol{v}_n$$

We want to observe the convergence behavior of the sequence of evaluations $\{p(\boldsymbol{x}_n)\}_{n\in\mathbb{N}}$ to $p(\boldsymbol{x}^\star)$. We have

$$\begin{aligned} p(\boldsymbol{x}_{n+1}) &= p(\boldsymbol{x}_n + \tfrac{\|\boldsymbol{v}_n\|^2}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)} \boldsymbol{v}_n) \\ &= \tfrac{1}{2}\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)\big(\tfrac{\|\boldsymbol{v}_n\|^2}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)}\big)^2 - \|\boldsymbol{v}_n\|^2 \tfrac{\|\boldsymbol{v}_n\|^2}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)} + p(\boldsymbol{x}_n) \\ &= p(\boldsymbol{x}_n) - \frac{\|\boldsymbol{v}_n\|^4}{2\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)}; \end{aligned}$$

therefore,

$$\begin{aligned} \frac{p(\boldsymbol{x}_{n+1}) - p(\boldsymbol{x}^\star)}{p(\boldsymbol{x}_n) - p(\boldsymbol{x}^\star)} &= \frac{p(\boldsymbol{x}_n) - p(\boldsymbol{x}^\star) - \frac{\|\boldsymbol{v}_n\|^4}{2\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)}}{p(\boldsymbol{x}_n) - p(\boldsymbol{x}^\star)} \\ &= 1 - \frac{\|\boldsymbol{v}_n\|^4}{2\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)\big(p(\boldsymbol{x}_n) - p(\boldsymbol{x}^\star)\big)} \\ &= 1 - \frac{\|\boldsymbol{v}_n\|^4}{2\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)\big(\frac{1}{2}\boldsymbol{x}_n Q \boldsymbol{x}_n^\intercal + D\boldsymbol{x}_n^\intercal + \frac{1}{2}DQ^{-1}D^\intercal\big)} \\ &= 1 - \frac{\|\boldsymbol{v}_n\|^4}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)\big(\boldsymbol{x}_n Q \boldsymbol{x}_n^\intercal + 2D\boldsymbol{x}_n^\intercal + DQ^{-1}D^\intercal\big)}. \end{aligned}$$

Note in the denominator we may rewrite some of the terms:

$$\begin{aligned} \boldsymbol{x}_n Q \boldsymbol{x}_n^\intercal &= \boldsymbol{x}_n Q (Q^{-1}Q) \boldsymbol{x}_n^\intercal = (\boldsymbol{x}_n Q) Q^{-1} (\boldsymbol{x}_n Q)^\intercal, \\ 2D\boldsymbol{x}_n^\intercal &= D\boldsymbol{x}_n^\intercal + D\boldsymbol{x}_n^\intercal = \boldsymbol{x}_n D^\intercal + D(Q^{-1}Q)\boldsymbol{x}_n^\intercal \\ &= \boldsymbol{x}_n (QQ^{-1}) D^\intercal + DQ^{-1}(\boldsymbol{x}_n Q)^\intercal \\ &= (\boldsymbol{x}_n Q) Q^{-1} D^\intercal + DQ^{-1}(\boldsymbol{x}_n Q)^\intercal. \end{aligned}$$

This allows us to rewrite in the following convenient form

$$\begin{aligned} \frac{p(\boldsymbol{x}_{n+1}) - p(\boldsymbol{x}^\star)}{p(\boldsymbol{x}_n) - p(\boldsymbol{x}^\star)} &= 1 - \frac{\|\boldsymbol{v}_n\|^4}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)(\boldsymbol{x}_n Q + D) Q^{-1} (\boldsymbol{x}_n Q + D)^\intercal} \\ &= 1 - \frac{\|\boldsymbol{v}_n\|^4}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)\mathcal{Q}_{(\boldsymbol{Q^{-1}})}(\boldsymbol{v}_n)}. \end{aligned}$$

We are ready to state the main result of this subsection:

THEOREM 3.8. *Given a d-dimensional vector D, and a positive definite symmetric matrix Q of size $d \times d$, consider the quadratic function $p(\boldsymbol{x}) = \frac{1}{2}\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{x}) + \langle D, \boldsymbol{x}\rangle$. Any sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ of steepest descent converges to the global minimum $\boldsymbol{x}^\star = -DQ^{-1}$. The sequence of evaluations $\{p(\boldsymbol{x}_n)\}_{n\in\mathbb{N}}$ converges linearly to $p(\boldsymbol{x}^\star) = -\frac{1}{2}\mathcal{Q}_{(\boldsymbol{Q^{-1}})}(D)$. In particular, if $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$ are the eigenvalues of Q, then*

$$\frac{p(\boldsymbol{x}_{n+1}) - p(\boldsymbol{x}^\star)}{p(\boldsymbol{x}_n) - p(\boldsymbol{x}^\star)} \leq \left(\frac{\lambda_d - \lambda_1}{\lambda_d + \lambda_1}\right)^2$$

PROOF. We start by offering the following lower bound estimate[3] involving the associated directions of steepest descent $\boldsymbol{v}_n$ in terms of the largest and smallest eigenvalues of Q. For all $n \in \mathbb{N}$,

$$\frac{\|\boldsymbol{v}_n\|^4}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)\mathcal{Q}_{(\boldsymbol{Q^{-1}})}(\boldsymbol{v}_n)} \geq \frac{4\lambda_0\lambda_d}{(\lambda_0 + \lambda_d)^2} \tag{10}$$

We have then

$$\begin{aligned}\frac{p(\boldsymbol{x}_{n+1}) - p(\boldsymbol{x}^\star)}{p(\boldsymbol{x}_n) - p(\boldsymbol{x}^\star)} &= 1 - \frac{\|\boldsymbol{v}_n\|^4}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)\mathcal{Q}_{(\boldsymbol{Q^{-1}})}(\boldsymbol{v}_n)} \\ &\leq 1 - \frac{4\lambda_1\lambda_d}{(\lambda_1 + \lambda_d)^2} = \left(\frac{\lambda_d - \lambda_1}{\lambda_d + \lambda_1}\right)^2 \qquad \square\end{aligned}$$

EXAMPLE 3.9. The global minimum value of the quadratic function $p(x, y) = 5x^2 + 5y^2 - xy - 11x + 11y + 11$ is zero, and found at $(1, -1)$. Notice that we may write this function in the form $p(x, y) = \frac{1}{2}\mathcal{Q}_{\boldsymbol{Q}}(x, y) + \langle D, [x, y]\rangle + 11$, where

$$\begin{aligned}D &= [-11, 11], \\ Q &= \begin{bmatrix} 10 & -1 \\ 10 & -1 \end{bmatrix}.\end{aligned}$$

The symmetric matrix Q has eigenvalues $\lambda_1 = 9 > 0$, $\lambda_2 = 11 > 0$ and is therefore positive definite. Theorem 3.8 states that sequences of steepest descent exhibit linear convergence with a rate of convergence not larger than $\delta = \left(\frac{11-9}{11+9}\right)^2 = 0.01$.

Observe the computations of the first six iterations for values of the ratios $\frac{p(\boldsymbol{x}_n)}{p(\boldsymbol{x}_{n-1})}$ when we use $(1.5, 3.5)$ as our initial guess.

[3]This is left as and advanced exercise. It is not too tricky; if you are stuck, see e.g. [**1**, section 1.3.1] for a proof.

n	x_n	y_n	$p(x_n, y_n)$	$\frac{p(x_n,y_n)}{p(x_{n-1},y_{n-1})}$
0	1.5000000000	3.5000000000	100.2500000000	
1	1.4498874016	−0.9600212545	1.0019989373	**0.0099950019**
2	1.0049975009	−0.9550224916	0.0100149812	**0.0099950019**
3	1.0044966254	−0.9996004124	0.0001000998	**0.0099950019**
4	1.0000499500	−0.9995504497	0.0000010005	**0.0099950019**
5	1.0000449438	−0.9999960061	0.0000000100	**0.0099950042**
6	1.0000004993	−0.9999955067	0.0000000001	**0.0099950528**

3. Broyden's Secant Method

Exercises

PROBLEM 3.1 (Basic). [**5**, p.249 #1] The following sequences all converge to zero.

$$v_n = n^{-10} \quad w_n = 10^{-n} \quad x_n = 10^{-n^2} \quad y_n = n^{10}3^{-n} \quad z_n = 10^{-3\cdot 2^n}$$

Indicate the type of convergence (See Appendix B).

PROBLEM 3.2 (Advanced). [**5**, p.249 #4] Give an example of a positive sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ converging to zero in such a way that $\lim_n \frac{\varepsilon_{n+1}}{\varepsilon_n^p} = 0$ for some $p > 1$, but not converging to zero with any order $q > p$.

PROBLEM 3.3 (Basic). Find an example of a function $f\colon \mathbb{R} \to \mathbb{R}$ (different from the function in Example 3.2) with a unique root at $x = 0$ for which the Newton-Raphson sequence is a loop no matter the initial guess $x_0 \neq 0$: $x_{2n} = x_0$, $x_{2n+1} = -x_0$ for all $n \in \mathbb{N}$. Bonus points is your function is trigonometric.

PROBLEM 3.4 (Intermediate). [**5**, p.251 #14] Consider the equation

$$x = \cos x.$$

(a) Show graphically that there exists a unique positive root $x^\star$. Indicate approximately where it is located.
(b) Show that Newton's method applied to $f(x) = x - \cos x$ converges for any initial guess $x_0 \in \left[0, \frac{\pi}{2}\right]$.

PROBLEM 3.5 (Intermediate). [**5**, p.251 #16] Consider the equation

$$\tan x + \lambda x = 0, \quad (0 < \lambda < 1).$$

(a) Show graphically, as simply as possible, that there is exactly one root $x^\star$ in the interval $\left[\frac{1}{2}\pi, \pi\right]$.
(b) Does Newton's method converge to the root $x^\star \in \left[\frac{1}{2}\pi, \pi\right]$ if the initial approximation is taken to be $x_0 = \pi$? Justify your answer.

PROBLEM 3.6 (Intermediate). [**5**, p.252 #17] Consider the equation

$$\log^2 x - x - 1 = 0, (x > 0).$$

(a) Graphical considerations suggest that there is exactly one positive solution $x^\star$, and that $0 < x^\star < 1$. Prove this.
(b) What is the largest positive $0 < x_0 \leq 1$ such that Newton's method with $f(x) = \log^2 x - x - 1$ started at x_0 converges to $x^\star$?

PROBLEM 3.7 (Advanced). [**5**, p.252 #18] Consider Kepler's equation

$$x - a\sin x - b = 0, \quad 0 < |a| < 1, \quad b \in \mathbb{R}$$

where a, b are parameters.

(a) Show that for each a, b there is exactly one real solution $x^\star = x^\star(a, b)$ that satisfies

$$b - |a| \leq x^\star(a, b) \leq b + |a|$$

(b) Let $m \in \mathbb{N}$ satisfy $m\pi < b < (m+1)\pi$. Show that Newton's method for $f(x) = x - a\sin x - b$ with starting value

$$x_0 = \begin{cases} (m+1)\pi & \text{if } (-1)^m a > 0 \\ m\pi & \text{otherwise} \end{cases}$$

is guaranteed to converge (monotonically) to $x^\star(a, b)$.

PROBLEM 3.8 (Basic). Consider the two equivalent equations

$$x\log x - 1 = 0, \tag{11}$$

$$\log x - \frac{1}{x} = 0. \tag{12}$$

(a) Show that there is exactly one positive root and find a rough interval containing it.
(b) For both (11) and (12), determine the largest interval on which Newton's method converges.
Hint: Investigate the convexity of the functions involved.

PROBLEM 3.9 (CAS). Design a process in `desmos.com` to test the search for critical points given by the recursion formulas produced by Newton's method.

PROBLEM 3.10 (CAS). The purpose of this problem is the design of *Horner's method* to evaluate polynomials effectively. Given a polynomial

$$p(x) = \sum_{k=0}^{n} a_k x^k = a_0 + a_1 x + \cdots + a_n x^n,$$

where $a_0, a_1, \ldots, a_n$ are real numbers, and given $x_0 \in \mathbb{R}$, we define the Horner's scheme $\{b_0, b_1, \ldots, b_n\}$ to evaluate $p(x_0)$ as follows:

$$\begin{aligned} b_n &= a_n \\ b_{n-1} &= a_{n-1} + b_n x_0 \\ b_{n-2} &= a_{n-2} + b_{n-1} x_0 \end{aligned}$$

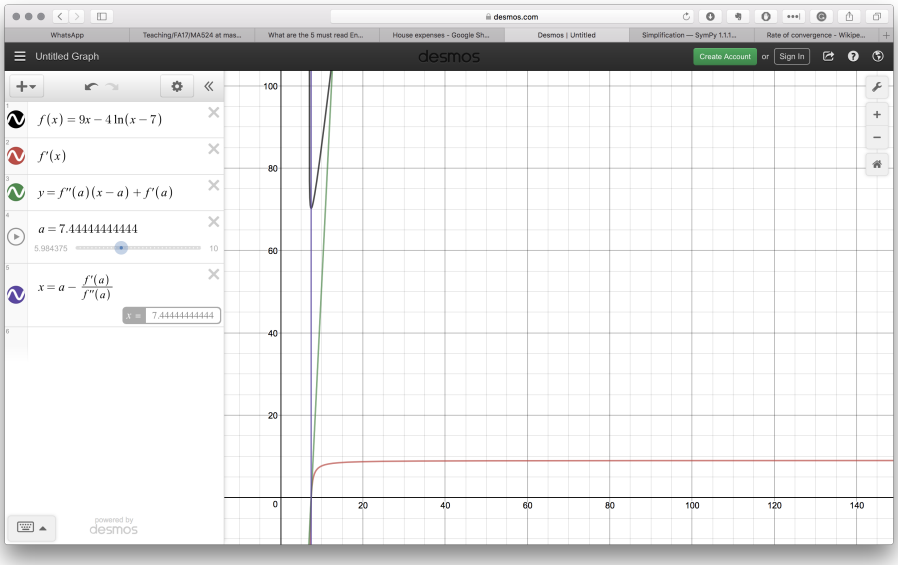

FIGURE 3.6. Newton method in `desmos.com`

$$\vdots$$

$$b_0 = a_0 + b_1 x_0$$

(a) Prove that $b_0 = p(x_0)$

(b) Use Horner's method to evaluate $p(x) = 2x^3 - 6x^2 + 2x - 1$ at $x = 3$. Illustrate all steps, and count the number of basic operations (addition, subtraction, multiplication, division) used.

(c) Employ the usual method of evaluation of polynomials to evaluate $p(x) = 2x^3 - 6x^2 + 2x - 1$ at $x = 3$. Count the number of basic operations (note that a raising to the cube counts as two multiplications, e.g.)

(d) In a computer language or CAS of your choice, write a routine to apply Horner's scheme to evaluate polynomials. Your routine should gather the following inputs:

- A list of coefficients `[a0, a1, ..., an]` representing the polynomial $p(x)$.
- A value x_0

The output of your routine should be $p(x_0)$.

PROBLEM 3.11 (CAS). In a computer language or CAS of your choice, design a routine that gathers the following as input:

- the definition of a generic real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$,
- the gradient ∇f of that function,
- an initial guess $\boldsymbol{x}_0 \in \mathbb{R}^d$,
- a number N of steps,

and produces the first $N+1$ terms of the Newton sequence to approximate a root of f.

Modify the previous routine to receive as input, instead of a number of steps, a *tolerance* `tol` indicating the accuracy of the solution. For example, if we require a root of the equation $f(x) = 0$ accurate to the first 16 correct decimal places, we use `tol = 1e-16`.

PROBLEM 3.12 (CAS). Use any of the routines that you wrote in Problem 3.11 to produce a table and a visual representation for the numerical solution of the following equations, with the given initial guesses and steps.

(a) $f(x) = \sin x$, with $x_0 = 0.5$, 5 steps.
(b) $f(x) = \sin x$, with $x = 3$, enough steps to obtain accurately the first 16 correct decimal places of π.
(c) $f(x) = -1 + \log x$, with $x = 2$, enough steps to obtain accurately the first 16 correct decimal places of e.

PROBLEM 3.13 (CAS). The objective of this problem is to use Newton's method to find an approximation to the golden ratio $\phi = \frac{1}{2}(1+\sqrt{5})$ accurate to the first 16 decimal places. Find first an appropriate polynomial $p(x)$ with integer coefficients for which ϕ is a root. Employ any of the routines that you wrote in Problem 3.11 with a good initial guess to guarantee the required result.

PROBLEM 3.14 (Intermediate—CAS). Consider the function

$$f(x) = 9x - 4\log(x-7).$$

We wish to study the behavior of Newton-Raphson to find approximations to the critical points of this function.

(a) Find the domain D of f.
(b) Find the global minimum of f analytically.
(c) Compute an exact formula for the Newton-Raphson iterate x_{n+1} for an initial guess $x_0 \in D$.
(d) Compute five iterations of the Newton-Raphson method starting at each of the following initial guesses:
    (a) $x_0 = 7.4$.
    (b) $x_0 = 7.2$.
    (c) $x_0 = 7.01$.
    (d) $x_0 = 7.8$.
    (e) $x_0 = 7.88$.
(e) Prove that the Newton-Raphson method converges to the optimal solution for any initial guess $x_0 \in (7, 7.8888)$.
(f) What is the behavior of the Newton-Raphson method if the initial guess is not in the interval $(7, 7.8888)$?

PROBLEM 3.15 (Intermediate—CAS). Consider the function

$$f(x) = 6x - 4\log(x-2) - 3\log(25-x).$$

We wish to study the behavior of Newton-Raphson to find approximations to the critical points of this function.

(a) Find the domain D of f.
(b) Find the global minimum of f analytically.
(c) Compute an exact formula for the Newton-Raphson iterate x_{n+1} for an initial guess $x_0 \in D$.
(d) Compute five iterations of the Newton-Raphson method starting at each of the following initial guesses:
    (a) $x_0 = 2.6$.
    (b) $x_0 = 2.7$.
    (c) $x_0 = 2.4$.
    (d) $x_0 = 2.8$.
    (e) $x_0 = 3$.
(e) Prove that the Newton-Raphson method converges to the optimal solution for any initial guess $x_0 \in (2, 3.05)$.
(f) What is the behavior of the Newton-Raphson method if the initial guess is not in the interval $(2, 3.05)$?

PROBLEM 3.16 (Advanced). [**1**, p.91 #1.4.1] The purpose of this exercise is to show that Newton's method is unaffected by linear scaling of the variables. Consider a linear invertible transformation of variables $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y}$. Write Newton's method in the space of the variables $\boldsymbol{y}$ and show that it generates the sequence $\boldsymbol{y}_n = \boldsymbol{A}^{-1}\boldsymbol{x}_n$, where $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ is the sequence generated by Newton's method in the space of variables $\boldsymbol{x}$.

PROBLEM 3.17 (Advanced). Prove Theorems 3.4, and 3.5.

PROBLEM 3.18 (Basic). Let $\boldsymbol{A}$ be a square matrix. An *LU-decomposition* is a factorization of $\boldsymbol{A} = \boldsymbol{L}\cdot\boldsymbol{U}$ into a *lower triangular* matrix $\boldsymbol{L}$ and an *upper triangular* matrix $\boldsymbol{U}$, both of which have non-zero entries in their diagonals. For example, the general case for 3×3 square matrices:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \underbrace{\begin{bmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{bmatrix}}_{\boldsymbol{L}} \cdot \underbrace{\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}}_{\boldsymbol{U}}$$

(a) Find an LU-decomposition of the following matrix

$$\begin{bmatrix} 4 & 3 \\ 6 & 3 \end{bmatrix}$$

that satisfies that all diagonal entries of $\boldsymbol{L}$ are ones.
(b) Find an example of a 2×2 square matrix for which there is not any possible LU-decomposition.

PROBLEM 3.19 (Advanced). Prove the following statements:

(a) A square matrix $\boldsymbol{A}$ of size $d \times d$ admits an LU-decomposition if and only if the leading principal minors are non-zero: $\Delta_k \neq 0$ for $1 \leq k \leq d$.

(b) If $\boldsymbol{A}$ is a symmetric positive definite matrix, then it is possible to find an LU-decomposition where $\boldsymbol{U} = \boldsymbol{L}^\intercal$: $\boldsymbol{A} = \boldsymbol{L} \cdot \boldsymbol{L}^\intercal$. In this case, this factorization is also called a *Cholesky decomposition*.

PROBLEM 3.20 (Advanced). We want to prove estimate (10) in the proof of Theorem 3.8 in page 39. This follows directly from the equivalent statement below, which is easier to handle. Prove the following result:

Kantorovich Estimate. Given a positive definite symmetric matrix Q of size $d \times d$, consider the quadratic function $p(\boldsymbol{x}) = \frac{1}{2}\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{x})$. Assume $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$ are the eigenvalues of Q. For any sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ of steepest descent for f, we have the following estimate involving the directions of steepest descent $\{\boldsymbol{v}_n\}_{n\in\mathbb{N}}$:

$$\frac{\|\boldsymbol{v}_n\|^4}{\mathcal{Q}_{\boldsymbol{Q}}(\boldsymbol{v}_n)\mathcal{Q}_{(\boldsymbol{Q}^{-1})}(\boldsymbol{v}_n)} \geq \frac{4\lambda_1\lambda_d}{(\lambda_1+\lambda_d)^2}$$

PROBLEM 3.21. Consider the quadratic polynomial

$$p(x, y) = \tfrac{1}{2}\mathcal{Q}_{\boldsymbol{Q}}(x, y) + \langle D, [x, y]\rangle + 13,$$

with

$$D = [4, -15]$$
$$Q = \begin{bmatrix} 10 & -9 \\ -9 & 10 \end{bmatrix}$$

(a) Find the global minimum value of p, and its location.
(b) Compute the eigenvalues of Q. Is Q positive definite?
(c) What is the worst-case scenario rate of convergence of sequences of steepest descent for this function?
(d) Compute sequences of steepest descent for this function with the initial guesses below. Make sure to report a table similar to the one in Example 3.9.
- $(0, 0)$
- $(-0.4, 0)$
- $(10, 0)$
- $(11, 0)$

PROBLEM 3.22. Consider the quadratic polynomial

$$p(x, y, z) = \tfrac{1}{2}\mathcal{Q}_{\boldsymbol{Q}}(x, y, z) + \langle D, [x, y, z]\rangle,$$

with

$$D = [12, -47, -8]$$
$$Q = \begin{bmatrix} 10 & -18 & 2 \\ -18 & 40 & -1 \\ 2 & -1 & 3 \end{bmatrix}$$

(a) Find the global minimum value of p, and its location.
(b) Compute the eigenvalues of Q. Is Q positive definite?
(c) What is the worst-case scenario rate of convergence of sequences of steepest descent for this function?
(d) Compute sequences of steepest descent for this function with the initial guesses below. Make sure to report a table similar to the one in Example 3.9.
    - $(0, 0, 0)$
    - $(15.09, 7.66, -6.56)$
    - $(11.77, 6.42, -4.28)$
    - $(4.46, 2.25, 1.85)$

CHAPTER 4

Existence and Characterization of Extrema for Constrained Optimization

CHAPTER 5

Numerical Approximation for Constrained Optimization

Index

Bibliography

[1] Dimitri P Bertsekas. *Nonlinear programming*. Athena scientific Belmont, 1999.

[2] Francisco J Blanco-Silva. *Mastering SciPy*. Packt Publishing Ltd, 2015.

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

[4] Robert Freund. Nonlinear programming. Massachusetts Institute of Technology: MIT OpenCourseWare, 2004. `https://ocw.mit.edu` License: Creative Commons BY-NC-SA.

[5] Walter Gautschi. *Numerical analysis*. Springer Science & Business Media, 2011.

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

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

[8] Walter Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964.

APPENDIX A

Basic sympy commands for Calculus

A typical **sympy** session usually starts by loading the *symbols* we need, some basic functions, and basic constructors. After that, we proceed to the description of the functions we require.

```
# Symbols, including one for infinity, π and e
from sympy.abc import x,y,t,h
from sympy import oo, pi, E
# Symbols with conditions
from sympy import var
a,b = var('a,b', positive=True)

# Basic functions we may need
from sympy import sqrt, sin, cos, tan, exp, log

# Some basic symbolic manipulations may be needed
from sympy import solve, factor, expand, simplify, limit

# To do vector calculus, we need these two as well
from sympy import Matrix
from sympy.tensor.array import derive_by_array

# If in a jupyter notebook, we may want to render output as LaTeX
from sympy import init_printing
init_printing()

# Description of f
f = sin(x)/x

# A generic Rosenbrock function
# Note the symbols a, b act as parameters, while x and y act as variables
R = (a-x)**2 + b*(y-x**2)**2
```

We are going to use these functions to perform several common operations in Calculus.

1. Function operations

Observe how easily we can perform all of the following:

Function evaluation: with the method

`.subs({variable1: value1, variable2: value2, ...})`

Limits: with the function `limit(object, variable, value)`.

Basic operations: with the usual operators for addition, subtraction, multiplication and division.

Composition: again with the method `.subs()`.

```
>>> f.subs({x: pi}) # f(π)
        0
>>> f.subs({x: 0})  # f(0) --- returns "not a number"
        nan
>>> limit(f, x, 0)  # Compute lim_{x→0} f(x) instead
        1
>>> (f.subs({x: x+h}) - f)/h  # A divided quotient...
        (sin(h + x)/(h + x) - sin(x)/x)/h
>>> limit( (f.subs({x: x+h}) - f)/h, h, 0) # ... and its limit as h → 0
        (x*cos(x) - sin(x))/x**2
```

Notice how smart **sympy** is in regard to the properties of symbols

```
>>> sqrt(x**2) # Square root of the square of a variable without conditions
        sqrt(x**2)
>>> sqrt(a**2) # Square root of the square of a positive variable
        a
```

Directional limits are also possible

```
>>> limit(1/x, x, 0, dir="+")
        oo
>>> limit(1/x, x, 0, dir="-")
        -oo
```

2. Derivatives, Gradients, Hessians

For functions of one variable, to obtain the symbolic derivative of a function (of any order), we usually employ the method

`.diff(variable, order)`

For functions of several variables, we employ instead

`derive_by_array(function, list-of-variables)`

If necessary, we may arrange our outputs as matrices, so we can employ proper matrix operations with them.

```
>>> f.diff(x) # f'(x) without the need to mess with limits
        cos(x)/x - sin(x)/x**2
>>> f.diff(x, 2)  # f''(x)
        (-sin(x) - 2*cos(x)/x + 2*sin(x)/x**2)/x
>>> derive_by_array(R, [x,y]) # The gradient of R, ∇R
        [-2*a - 4*b*x*(-x**2 + y) + 2*x, b*(-2*x**2 + 2*y)]
>>> gradient = _   # Store that in the variable 'gradient'
>>> derive_by_array(gradient, [x,y]) # The Hessian of R, HessR
        [[8*b*x**2 - 4*b*(-x**2 + y) + 2, -4*b*x], [-4*b*x, 2*b]]
>>> hessian = Matrix(2,2, _) # Store that as a matrix, call it 'hessian'
>>> hessian[0,0] # If we want to access the first entry of the matrix
        8*b*x**2 - 4*b*(-x**2 + y) + 2
>>> simplify(_) # Simplify that expression
```

```
        12*b*x**2 - 4*b*y + 2
>>> Delta1 = _  # Store that value as 'Delta1'
>>> hessian.det()  # Compute the determinant of the Hessian
        -16*b**2*x**2 + 2*b*(8*b*x**2 - 4*b*(-x**2 + y) + 2)
>>> Delta2 = simplify(_) # Store that value as 'Delta2'
```

It is then a simple task (in some cases) to search for critical points by solving symbolically $\nabla f = 0$, and checking whether they are local maxima, local minima or saddle points.

```
>>> solve(gradient, [x,y]) # Critical points of R
        [(a, a**2)]
>>> crit_points = _ # This is a list.  We call it 'crit_points'
>>> for point in crit_points:
...     x0,y0 = point
...     print(point)
...     print("Delta1 = ", Delta1.subs({x:x0, y:y0}))
...     print("Delta2 = ", Delta2.subs({x:x0, y:y0}))
...
        (a, a**2)
        Delta1 =  8*a**2*b + 2
        Delta2 =  4*b
>>> 8*a**2*b + 2 > 0  # Is Delta1 > 0? (remember a,b>0)
        True
>>> 4*b > 0 # Is Delta2 > 0?
        True
```

The conclusion after this small session is that any Rosenbrock function $R(x, y) = (a - x)^2 + b(y - x^2)^2$ has a global minimum at the point (a, a^2).

A word of warning. Symbolic differentiation and manipulation of expressions may not work in certain cases. For those, numerical approximation is more suited (and incidentally, *that* is the reason you are taking this course).

```
>>> solve(f.diff(x))
NotImplementedError: multiple generators [x, tan(x/2)]
No algorithms are implemented to solve equation
x**2*(-tan(x/2)**2 + 1)/(tan(x/2)**2 + 1) - 2*x*tan(x/2)/(tan(x/2)**2 + 1)
```

3. Integration

Symbolic integration for the computation of antiderivatives is also possible. Definite integrals, while the symbolic setting allows it in many cases, it is preferably done in a numerical setting.

```
>>> R.integrate(x) # ∫ R(x, y) dx
        -a*x**2 + b*x**5/5 + x**3*(-2*b*y/3 + 1/3) + x*(a**2 + b*y**2)
>>> R.integrate(y) # ∫ R(x, y) dy
        -b*x**2*y**2 + b*y**3/3 + y*(a**2 - 2*a*x + b*x**4 + x**2)
>>> R.integrate(x, (x, 0, 1)).integrate(y, (y, 0, 1)) # ∫_0^1 ∫_0^1 R(x, y) dx dy
        a**2/4 - a/6 + 11*b/360 + 1/24
>>> f.integrate(x) # ∫ sin(x)/x dx
        Si(x)
```

```
>>> f.integrate(x, (x, 0, pi)) # ∫_0^π sin(x)/x dx
        -2 + pi*Si(pi)
>>> _.evalf() # How much is that, actually?
        3.81803183741885
```

4. Sequences, series

```
>>>
```

5. Power series, series expansions

```
>>>
```

APPENDIX B

Rates of Convergence

DEFINITION. Consider a convergent sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^d$ with $\boldsymbol{x}^\star = \lim_n \boldsymbol{x}_n$. We say that this sequence exhibits

Linear Convergence: If there exists $0 < \delta < 1$ so that

$$\lim_n \frac{\|\boldsymbol{x}_{n+1} - \boldsymbol{x}^\star\|}{\|\boldsymbol{x}_n - \boldsymbol{x}^\star\|} = \delta.$$

We refer to δ as the *rate of convergence*.

Superlinear Convergence: If

$$\lim_n \frac{\|\boldsymbol{x}_{n+1} - \boldsymbol{x}^\star\|}{\|\boldsymbol{x}_n - \boldsymbol{x}^\star\|} = 0.$$

Sublinear Convergence: If

$$\lim_n \frac{\|\boldsymbol{x}_{n+1} - \boldsymbol{x}^\star\|}{\|\boldsymbol{x}_n - \boldsymbol{x}^\star\|} = 1.$$

If, additionally,

$$\lim_n \frac{\|\boldsymbol{x}_{n+2} - \boldsymbol{x}_{n+1}\|}{\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\|} = 1,$$

we say the the sequence exhibits *logarithmic convergence* to $\boldsymbol{x}^\star$.

Convergence of order $q > 1$: If $\boldsymbol{x}_n$ is exhibits superlinear convergence, and there exists $q > 1$, $0 < \delta < 1$ so that

$$\lim_n \frac{\|\boldsymbol{x}_{n+1} - \boldsymbol{x}^\star\|}{\|\boldsymbol{x}_n - \boldsymbol{x}^\star\|^q} = \delta.$$

In particular,

- Convergence with $q = 2$ is said to be *quadratic*.
- Convergence with $q = 2$ is said to be *cubic*.
- etc.

A practical method to calculate the rate of convergence of a sequence is to calculate the following sequence, which converges to q:

$$q \approx \frac{\log\left|\frac{x_{n+1}-x_n}{x_n-x_{n-1}}\right|}{\log\left|\frac{x_n-x_{n-1}}{x_{n-1}-x_{n-2}}\right|} \tag{13}$$

EXAMPLE B.1. The sequence $x_n = 1/n!$ exhibits superlinear convergence, since $\lim_n \frac{1}{n!} = 0$ and

$$\lim_n \frac{x_{n+1}}{x_n} = \lim_n \frac{1}{n+1} = 0.$$

EXAMPLE B.2. Given $a \in \mathbb{R}$, $0 < r < 1$, the geometric sequence $\boldsymbol{x}_n = ar^n$ exhibits linear convergence, since $\lim_n ar^n = 0$ and

$$\lim_n \frac{x_{n+1}}{x_n} = r < 1.$$

The rate of convergence is precisely r.

EXAMPLE B.3. The sequence $x_n = 2^{-2^n}$ converges to zero and is superlinear:

$$\lim_n \frac{x_{n+1}}{x_n} = \lim_n 2^{-2^n} = 0$$

Using the estimation for q given by the formula in (13), we obtain that this sequence exhibits quadratic convergence.

EXAMPLE B.4. The sequence $x_n = 1/n$ converges to zero and is sublinear, since

$$\lim_n \frac{x_{n+1}}{x_n} = \lim_n \frac{n}{n+1} = 1.$$

Notice

$$\lim_n \frac{|\boldsymbol{x}_{n+2} - \boldsymbol{x}_{n+1}|}{|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n|} = \lim_n \frac{n}{n+2} = 1;$$

therefore, this sequence exhibits logarithmic convergence.