ISyE 6663 – Nonlinear Optimization

Johannes Milz

Fall 2023
Last update: December 6, 2023

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology, Atlanta, GA

The lecture notes are mainly based on

- A. Ben-Tal and A. Nemirovski: Optimization III: Convex Analysis, Nonlinear Programming Theory, Standard Nonlinear Programming Algorithms, lecture notes, Georgia Tech, 2023, `https://www2.isye.gatech.edu/~nemirovs/`
- G. Lan: Lectures for Convex and Nonlinear Optimization, lecture notes, Georgia Tech, 2022

Contents

1 Convex Sets

1.1 Definition and Examples

We begin by defining convex sets.

Definition 1.1. *A set $X \subset \mathbb{R}^n$ is said to be* convex *if*

$$\lambda x + (1-\lambda)y \in X \quad \textit{for all} \quad x, y \in X, \quad \lambda \in [0,1].$$

For x, $y \in X$, the set of all points $\lambda x + (1-\lambda)y$ for $\lambda \in [0,1]$ is called *line segment*. Hence, a set $X \subset \mathbb{R}^n$ is convex if the line segment of any pairs of points from X is contained in X. Figure 1.1 depicts two convex and one nonconvex set.

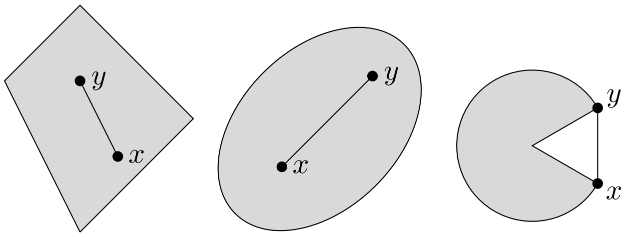

Figure 1.1: Two convex sets and one nonconvex set. The set on the right is nonconvex.

For $\lambda \in [0,1]$ and x, $y \in X$, the point $\lambda x + (1-\lambda)y$ is called a *convex combination* of x and y.

We provide simple examples of convex sets.

Example 1.2.

1. The Euclidean space $\mathbb{R}^n$ is a convex set.
2. The empty set $\emptyset$ is convex.
3. For each $x \in \mathbb{R}^n$, the singleton $\{x\}$ is a convex set.
4. The solution set of an arbitrary (finite or infinite) system of linear inequalities,

   $$X = \{x \in \mathbb{R}^n \colon a_\alpha^T x \le b_\alpha, \quad \alpha \in \mathcal{A}\},$$

   is convex, where $a_\alpha \in \mathbb{R}^n$ and $b_\alpha \in \mathbb{R}$, and $\mathcal{A}$ is some (index) set. See Figure 1.5 for an example.
   To establish this assertion, let x, $y \in X$ and let $\lambda \in [0,1]$. We have

   $$a_\alpha^T x \le b_\alpha, \quad a_\alpha^T y \le b_\alpha, \quad \text{for all} \quad \alpha \in \mathcal{A}.$$

Hence for all $\alpha \in \mathcal{A}$,

$$\begin{aligned} a_\alpha^T[\lambda x + (1-\lambda)y] &= \lambda a_\alpha^T x + (1-\lambda)a_\alpha^T y \\ &\leq \lambda b_\alpha + (1-\lambda)b_\alpha \\ &= b. \end{aligned}$$

We obtain $\lambda x + (1-\lambda)y \in X$.

Linear Subspaces

A set $L \subset \mathbb{R}^n$ is called a *linear subspace* of $\mathbb{R}^n$ if it is nonempty and $\lambda x + \mu y \in L$ for all x, $y \in L$ and λ, $\mu \in \mathbb{R}$. Each linear subspace L of $\mathbb{R}^n$ contains the zero vector $0 = (0, \ldots, 0) \in \mathbb{R}^n$, as L is nonempty and hence $0 \cdot x + 0 \cdot x \in L$ for all $x \in L$. For example, the set $L = \{x \in \mathbb{R}^2 \colon (-1,1)^T x = 0\}$ is a linear subspace of $\mathbb{R}^2$; see Figure 1.2.

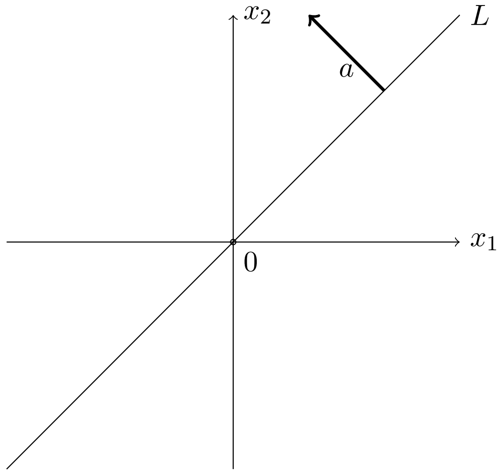

Figure 1.2: Linear subspace $L = \{x \in \mathbb{R}^2 \colon a^T x = 0\}$ in $\mathbb{R}^2$ with $a = (-1,1) \in \mathbb{R}^2$. The coordinate system's origin is 0. A point $x \in \mathbb{R}^2$ is in L if and only if it is orthogonal to a.

The smallest linear subspace of $\mathbb{R}^n$ is the singleton $\{0\}$. Linear subspaces are convex sets, as the definitions of linear subspaces and convex sets show.

Linear subspaces can be constructed through other sets. For a nonempty set $X \subset \mathbb{R}^n$, we call all vectors of the form $\sum_{i=1}^m \lambda_i x^i$ with $m \in \mathbb{N}$, $\lambda_i \in \mathbb{R}$, and $x^i \in X$ for $i = 1, \ldots, m$ a *linear combination* of the points $x^1, \ldots, x^m$. Moreover, we denote by $\mathrm{Lin}(X)$ the set of all *finite* linear combinations of elements of X. The set $\mathrm{Lin}(X)$ is called *linear span* of X. It is the smallest linear subspace containing X, that is, if L is a linear subspace with $X \subset L$, then $\mathrm{Lin}(X) \subset L$. In particular, if L is a linear subspace of $\mathbb{R}^n$, then we have $\mathrm{Lin}(L) = L$.

Let $x^1, \ldots, x^m$ be a collection of m vectors in $\mathbb{R}^n$. The collection is called *linearly* *independent* if no nontrival combination of these points is zero, that is, if

$$\sum_{i=1}^{m} \lambda_i x^i = 0 \quad \text{implies} \quad \lambda_i = 0, \quad i = 0, \ldots, m.$$

Each linear subspace of $\mathbb{R}^n$ has a basis. The (linear) *dimension* of a linear subspace L is the number of elements in a basis of L. A linear subspace L of $\mathbb{R}^n$ can be represented by the solution set of finitely many homogeneous linear equations, that is, $L = \{x \in \mathbb{R}^n \colon a_i^T x = 0,\ i = 1, \ldots, m\}$ for some $a_i \in \mathbb{R}^n$ and $m \in \mathbb{N}$. Let $A \in \mathbb{R}^{m \times n}$ be a matrix such that its ith row is equal to a_i^T. Using this definition, we can write $L = \{x \in \mathbb{R}^n \colon Ax = 0\}$. Let us provide details for these assertions. Let $L \subset \mathbb{R}^n$ be a linear subspace with dimension $r \in \mathbb{N}$. Then there exists a basis of $m := n - r$ vectors $a_1, \ldots, a_m$ of the linear subspace $L^\perp := \{y \in \mathbb{R}^n \colon y^T x = 0 \text{ for all } x \in L\}$, the orthogonal complement of L. We let a_i^T be the rows of a matrix $A \in \mathbb{R}^{m \times n}$. The matrix A has rank $m = n - r$ and we have $Ax = 0$ for all $x \in L$. Since L has dimension r and $\{x \in \mathbb{R}^n \colon Ax = 0\}$ has dimension $n - m = n - (n - r) = r$, we obtain $L = \{x \in \mathbb{R}^n \colon Ax = 0\}$. Alternatively, we can write $L = \{x \in \mathbb{R}^n \colon \exists y \in \mathbb{R}^r \text{ with } x = By\}$, where the columns of $B \in \mathbb{R}^{n \times r}$ consist of a basis of L. We have $AB = 0$.

Affine Subspaces

A set $M \subset \mathbb{R}^n$ is called an *affine subspace* if $\lambda x + \mu y \in M$ for all x, $y \in M$ and λ, $\mu \in \mathbb{R}$ with $\lambda + \mu = 1$. For $x^i \in M$ and $\lambda_i \in \mathbb{R}$ with $\sum_{i=1}^m \lambda_i = 1$, the vector $\sum_{i=1}^m \lambda_i x^i$ is called *affine combination* of the points $x^1, \ldots, x^m$. Let $X \subset \mathbb{R}^n$ be a nonempty set. We denote by by $\operatorname{Aff}(X)$ the set of all *finite* affine combinations of elements of X. The set $\operatorname{Aff}(X)$ is called *affine hull* of X. The affine hull of X is the smallest affine subspace containing X.

A nonempty set $M \subset \mathbb{R}^n$ is an affine subspace of $\mathbb{R}^n$ if and only if there exists $a \in M$ such that $L := M - a := \{y = x - a \colon x \in M\}$ is a linear subspace of $\mathbb{R}^n$. In this case, we have $a \in M$ and $L = M - b$ for all $b \in M$. Moreover, an affine subspace M is a linear subspace if and only if $0 \in M$. The *affine dimension* of M is the linear dimension of the linear subspace L.

These facts ensure that a nonempty affine subspace M of $\mathbb{R}^n$ is a set given by $M = a + L := \{a + x \colon x \in L\}$, where $L \subset \mathbb{R}^n$ is a linear subspace of $\mathbb{R}^n$ and $a \in \mathbb{R}^n$ is a vector. Geometrically, affine subspaces are shifts of linear subspaces by some vector. As discussed above, L can be represented as the solution set to finitely many homogeneous linear questions. As a result, M can be written as the solution set of finitely many nonhomogeneous linear equations, that is, $M = \{x \in \mathbb{R}^n \colon a_i^T x = a^T a_i,\ i = 1, \ldots, m\}$, where $a_i \in \mathbb{R}^n$, $i = 1, \ldots, m$ are the vectors that define the linear subspace L. Figures 1.3 and 1.4 provide illustrations.

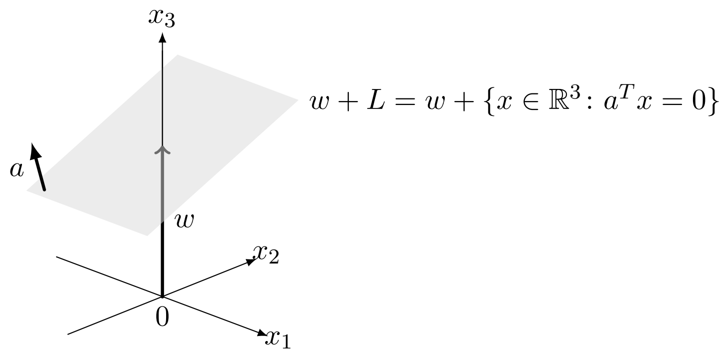

Figure 1.3: An affine subspace M in $\mathbb{R}^3$ given by $w + L$ for some linear subspace $L \subset \mathbb{R}^3$. The affine subspace M is obtained by shifting L by the vector a. Here $w = (0, 0, 1)$ and $L = \{x \in \mathbb{R}^3 \colon a^T x = 0\}$. The coordinate systems's origin is 0.

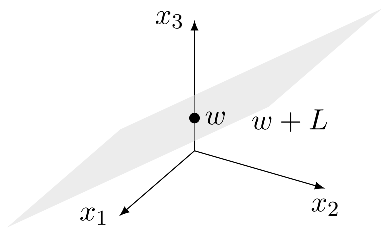

Figure 1.4: An affine subspace M in $\mathbb{R}^3$ given by $a + L$ for some linear subspace $L \subset \mathbb{R}^3$. The affine subspace M is obtained by shifting L by the vector a.

For any set $X \subset \mathbb{R}^n$ with $a \in X$, we have

$$\operatorname{Aff}(X) = a + \operatorname{Lin}(X - a).$$

Let $x^0, \ldots, x^m$ be a collection of $m + 1$ vectors in $\mathbb{R}^n$. The collection is called *affinely independent* if no nontrival combination of these points with zero sum coefficients is zero, that is, if

$$\sum_{i=0}^{m} \lambda_i x^i = 0 \quad \text{and} \quad \sum_{i=0}^{m} \lambda_i = 0 \quad \text{implies} \quad \lambda_i = 0, \quad i = 0, \ldots, m.$$

We find that $x^0, \ldots, x^m$ are affinely independent if and only if the vectors $(x^i, 1) \in \mathbb{R}^{n+1}$, $i = 0, \ldots, m$ are linearly independent. Here, $(x^i, 1)$ is the vector x^i augmented by adding the component 1 to it. Moreover, the vectors $x^0, \ldots, x^m$ are affinely independent if and only if $x^1 - x^0, \ldots, x^m - x^0$ is linearly independent.

Let $M \subset \mathbb{R}^n$ be a nonempty affine subspace of dimension m. Then we can write $M = a + L = a + \{x \in \mathbb{R}^n \colon Ax = 0\}$ for some matrix $A \in \mathbb{R}^{n-m \times n}$ and a vector $a \in \mathbb{R}^n$. Alternatively, we can write $M = \{x \in \mathbb{R}^n \colon \exists y \in \mathbb{R}^m \text{ with } x = By - a\}$, where $B \in \mathbb{R}^{n \times m}$ with $AB = 0$.

Polyhedral Sets

A set $X \subset \mathbb{R}^n$ is called *polyhedral* if it takes the form

$$X = \{x \in \mathbb{R}^n \colon Ax \leq b\},$$

for some matrix $A \in \mathbb{R}^{m \times n}$, a vector $b \in \mathbb{R}^m$, and a natural number $m \in \mathbb{N}$. If a_i^T is the ith row of the matrix A, we may also write

$$X = \{x \in \mathbb{R}^n \colon Ax \leq b\} = \{x \in \mathbb{R}^n \colon a_i^T x \leq b_i, \quad i = 1, \ldots, m\}.$$

In other words, a set is polyhedral if it is the solution set of a finite system of nonstrict linear inequalities. Figure 1.5 depicts a polyhedral set.

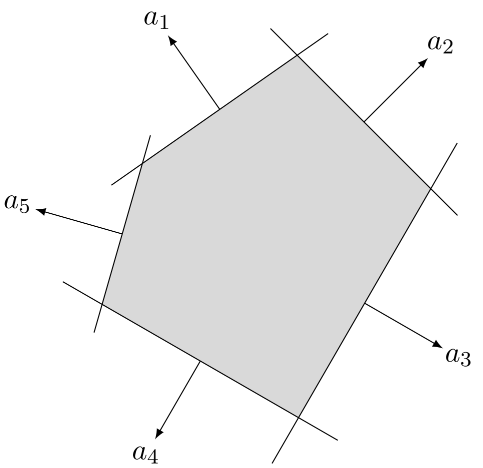

Figure 1.5: A polyhedral set given by five nonstrict linear inequalities.

A bounded nonempty polyhedral set is called *polytope*. Figures 1.5 and 1.6 depict polytopes.

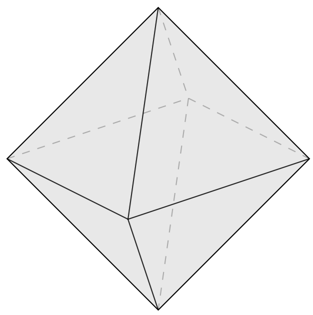

Figure 1.6: A polytope.

Polyhedral sets are convex. To verify this assertion, let $X = \{x \in \mathbb{R}^n \colon Ax \leq b\}$. Let x, $y \in X$ and let $\lambda \in [0,1]$. We have $A[\lambda x + (1-\lambda)y] = \lambda Ax + (1-\lambda)Ay \leq \lambda b + (1-\lambda)b = b$. Hence $\lambda x + (1-\lambda)y \in X$. As a result, polyhedral sets are convex.

Norm balls

For each norm $\|\cdot\|$ on $\mathbb{R}^n$, the *unit norm ball* $\{x \in \mathbb{R}^n \colon \|x\| \le 1\}$ is convex. Before, we establish the convexity of unit norm balls, we recall the definition of norms. A function $\|\cdot\| \colon \mathbb{R}^n \to \mathbb{R}$ is called a *norm* on $\mathbb{R}^n$ if (i) $\|x\| \ge 0$ for all $x \in \mathbb{R}^n$, and $\|x\| = 0$ if and only if $x = 0$, (ii) $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^n$, and (iii) $\|x+y\| \le \|x\| + \|y\|$ for all x, $y \in \mathbb{R}^n$. The latter property is referred to as *triangle inequality.* We provide three examples of norms on $\mathbb{R}^n$: (i) the one-/1-norm $\|x\|_1 \coloneqq \sum_{i=1}^n |x_i|$, (ii) the Euclidean norm $\|x\|_2 \coloneqq \sqrt{\sum_{i=1}^n x_i^2}$, and (iii) the uniform norm $\|x\|_\infty \coloneqq \max_{1 \le i \le n} |x_i|$. Figure 1.7 depicts graphs of their unit norm balls.

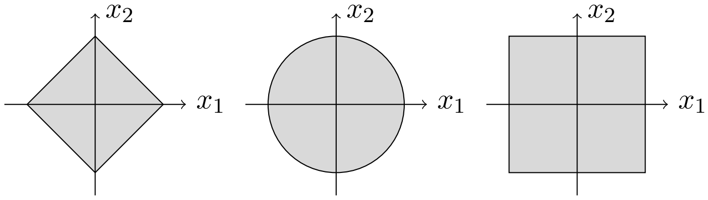

Figure 1.7: Unit norm balls of 1-norm (left), Euclidean norm, and uniform norm (right).

For a norm $\|\cdot\|$ on $\mathbb{R}^n$, a scalar $r \ge 0$ and a vector $a \in \mathbb{R}^n$, we show that the norm ball $\{x \in \mathbb{R}^n \colon \|x - a\| \le r\}$ with center a and radius r is convex. Let x, $y \in \mathbb{R}^n$ with $\|x - a\| \le r$ and $\|y - a\| \le r$ and let $\lambda \in [0, 1]$. We compute

$$\begin{aligned}
\|\lambda x + (1-\lambda)y - a\| &= \|\lambda(x-a) + (1-\lambda)(y-a)\| \\
&\le \|\lambda(x-a)\| + \|(1-\lambda)(y-a)\| \\
&= \lambda\|x-a\| + (1-\lambda)\|y-a\| \\
&\le \lambda r + (1-\lambda) r.
\end{aligned}$$

Hence the set $\{x \in \mathbb{R}^n \colon \|x - a\| \le r\}$ is convex. Using similar arguments, we can also show that the set $\{x \in \mathbb{R}^n \colon \|x - a\| < r\}$ is convex.

Ellipsoids

A set X is called an *ellipsoid* if it can be represented by

$$X = \{x \in \mathbb{R}^n \colon (x-a)^T Q (x-a) \le r^2\},$$

where $Q \in \mathbb{R}^{n \times n}$ is symmetric positive definite, $a \in \mathbb{R}^n$ is a vector (called the ellipsoid's center), and a scalar $r > 0$ (called the ellipsoid's radius). Figure 1.8 depicts an ellipsoid.

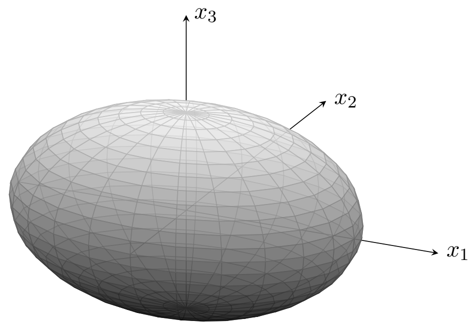

Figure 1.8: An ellipsoid in $\mathbb{R}^3$.

We recall some terminology and facts from linear algebra. A matrix $A \in \mathbb{R}^{n\times n}$ is called a *square* matrix. A matrix $A \in \mathbb{R}^{n\times n}$ is *positive definite* if $d^T Ad > 0$ for all $d \in \mathbb{R}^n \setminus \{0\}$. Finally, $A \in \mathbb{R}^{n\times n}$ is referred to as *symmetric* if $A = A^T$, where A^T is the transpose of A. A symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$ has a unique "square root" $A^{1/2} \in \mathbb{R}^{n\times n}$ with $A = A^{1/2}A^{1/2}$ and $A^{1/2}$ is symmetric positive definite.

Let us demonstrate that ellipsoids are convex. Let X be the ellipsoid introduced above. The matrix Q is positive definite and symmetric and hence as a symmetric positive definite square root $Q^{1/2}$ with $Q = Q^{1/2}Q^{1/2}$. The function $\|x\|_Q = \|Q^{1/2}x\|_2$ defines a norm on $\mathbb{R}^n$, as $Q^{1/2}$ is symmetric positive definite. We obtain

$$\begin{aligned}\{x \in \mathbb{R}^n \colon (x-a)^T Q(x-a) \le r^2\} &= \{x \in \mathbb{R}^n \colon \|x-a\|_Q^2 \le r^2\}\\ &= \{x \in \mathbb{R}^n \colon \|x-a\|_Q \le r\}.\end{aligned}$$

The latter set is a norm ball with center a and radius r and as such convex.

ε-Neighborhoods of Convex Sets

If $M \subset \mathbb{R}^n$ is a nonempty convex set, $\|\cdot\|$ is a norm on $\mathbb{R}^n$, and $\varepsilon \in [0,\infty)$, then the *$\varepsilon$-neighborhood* of M,

$$X = \{x \in \mathbb{R}^n \colon \inf_{y\in M} \|x-y\| \le \varepsilon\},$$

is convex. For $x \in \mathbb{R}^n$, we define the distance from x to M by $\operatorname{dist}_{\|\cdot\|}(x, M) := \inf_{y\in M} \|x-y\|$.

To show this convexity statement, we observe at $x \in X$ if and only if for each $\varepsilon' > \varepsilon$, there exists $y \in M$ such that $\|x-y\| \le \varepsilon'$. (This fact is a consequence of the definition of the infimum and $\varepsilon < \infty$.) Now, let x, $y \in X$ and let $\lambda \in [0,1]$. We fix $\varepsilon' > \varepsilon$. The above statement implies that there exist u, $v \in M$ with $\|x-u\| \le \varepsilon'$ and $\|y-v\| \le \varepsilon'$. Defining $w = \lambda u + (1-\lambda)v$, we find that $w \in M$

$$\begin{aligned}\|\lambda x + (1-\lambda)y - w\| &= \|\lambda(x-u) + (1-\lambda)(y-v)\|\\ &\le \lambda\|x-u\| + (1-\lambda)\|y-v\| \le \varepsilon'.\end{aligned}$$

Putting together the pieces, we find that X is a convex set.

1.2 Convex Combinations and Convex Hulls

Let $x^1, \ldots, x^m \in \mathbb{R}^n$ be vectors. If $m \in \mathbb{N}$, $\sum_{i=1}^m \lambda_i = 1$ and $\lambda_i \geq 0$, then the vector $\sum_{i=1}^m \lambda_i x^i$ is called a *convex combination* of the points $x^1, \ldots, x^m$. In the definition of convex sets, we have encountered convex combinations with $m = 2$. Figure 1.9 depicts a convex combination of three points in $\mathbb{R}^3$.

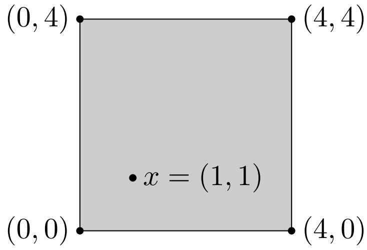

Figure 1.9: $x = (1, 1)$ is a convex combination of $(0, 0)$, $(4, 0)$, and $(0, 4)$.

Using the notion of convex combinations, we obtain an equivalent characterization of convexity of a set.

Lemma 1.3. *A set $X \subset \mathbb{R}^n$ is convex if and only if each convex combination of vectors from X is contained in X.*

Proof. To establish this equivalent characterization, we demonstrate both implications. Let X be convex. Let us prove that every $m$term convex combination of vectors from X is contained in X using induction. If $m = 1$ and $m = 2$, then the statement holds true. Suppose that the statement is true for m. Let $x^1, \ldots, x^{m+1} \in X$ and $\lambda_i \geq 0$ with $\sum_{i=1}^{m+1} \lambda_i = 1$. If $\lambda_{m+1} = 1$, then $\sum_{i=1}^{m+1} \lambda_i x^i = x^{m+1} \in X$. If $\lambda_{m+1} < 1$, then

$$\sum_{i=1}^{m+1} \lambda_i x^i = (1 - \lambda_{m+1}) \underbrace{\left(\sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} x^i \right)}_{\in X} + \lambda_{m+1} x_{m+1}.$$

Since X is convex, we obtain $\sum_{i=1}^{m+1} \lambda_i x^i \in X$.

To establish the reverse direction, we choose $m = 2$ in the definition of convex combinations. $\square$

The intersection $\cap_{\alpha \in \mathcal{A}} X_\alpha$ of an arbitrary family of convex set $(X_\alpha)_{\alpha \in \mathcal{A}}$ is a convex set. Here, $\mathcal{A}$ is some set. Let $X \subset \mathbb{R}^n$ be a set. The intersection of all convex sets containing X is called *convex hull* of X and is denoted by $\operatorname{Conv}(X)$. In other words, the convex hull of a set X is the smallest convex set containing X. Figure 1.10 depicts convex hull of points in $\mathbb{R}^2$.

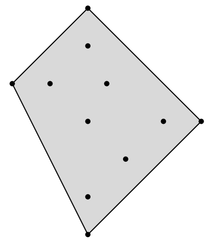

Figure 1.10: Convex hull of points.

The convex hull has an explicit representation in terms of convex combinations of elements of X. We show that the convex hull $\operatorname{Conv}(X)$ of a set $X \subset \mathbb{R}^n$ is the set of all convex combinations of points from X.

Proposition 1.4. *If $X \subset \mathbb{R}^n$ is a set, then*

$$\operatorname{Conv}(X) = \Big\{ \sum_{i=1}^{m} \lambda_i x^i \colon\ m \in \mathbb{N},\, x^i \in X,\, \lambda_i \geq 0,\, \sum_{i=1}^{m} \lambda_i = 1 \Big\},$$

Proof. Let us define the set $\widehat{X} := \{ \sum_{i=1}^{m} \lambda_i x^i \colon\ m \in \mathbb{N},\, x^i \in X,\, \lambda_i \geq 0,\, \sum_{i=1}^{m} \lambda_i = 1 \}$.

Since $X \subset \operatorname{Conv}(X)$ and $\operatorname{Conv}(X)$ is convex, Lemma 1.3 ensures that each convex combination of points from X is contained in $\operatorname{Conv}(X)$. Hence $\widehat{X} \subset \operatorname{Conv}(X)$.

Let $x,\, y \in \widehat{X}$. Fix $\nu \in [0,1]$. We show that $\nu x + (1-\nu) y \in \widehat{X}$. Since $x,\, y \in \widehat{X}$, $x = \sum_{i=1}^{m} \lambda_i x^i$ and $y = \sum_{j=1}^{p} \mu_j y^j$ for some $x^i \in X$, $y^j \in X$, $\lambda_i \geq 0$, $\mu_j \geq 0$ and $\sum_{i=1}^{m} \lambda_i = 1$ and $\sum_{j=1}^{p} \mu_j = 1$. Moreover,

$$\nu \sum_{i=1}^{m} \lambda_i + (1-\nu) \sum_{j=1}^{p} \mu_j = \nu + (1-\nu) = 1,$$

$\nu \lambda_i \geq 0$ and $(1-\nu)\mu_j \geq 0$. We obtain

$$\nu x + (1-\nu) y = \sum_{i=1}^{m} \nu \lambda_i x^i + \sum_{j=1}^{p} (1-\nu) \mu_j y^j \in \widehat{X}.$$

Hence $\widehat{X}$ is convex. Combined with $X \subset \widehat{X}$ and the fact that $\operatorname{Conv}(X)$ is the smallest convex set containing X, we have $\operatorname{Conv}(X) \subset \widehat{X}$. □

Simplices

Let $x^0, \ldots, x^m$ be a collection of $m+1$ affinely independent vectors in $\mathbb{R}^n$. An m-dimensional *simplex* Δ with vertices $x^0, \ldots, x^m$ is the convex hull of the affinely independent points $x^0, \ldots, x^m$ in $\mathbb{R}^n$, that is,

$$\Delta = \Delta(x^0, \ldots, x^m) := \operatorname{Conv}(\{x^0, \ldots, x^m\}).$$

Let $e^1, \ldots, e^n$ be the standard/canonical basis vectors in $\mathbb{R}^n$. For $1 \leq i \leq n$, the ith entry of e^i is one and all others are zero. The corresponding $(n-1)$-dimensional simplex is the *standard simplex*, the convex hull of $e^1, \ldots, e^n$. We have

$$\operatorname{Conv}(\{e^1, \ldots, e^n\}) = \left\{ x \in \mathbb{R}^n \colon x_i \geq 0, \sum_{i=1}^{n} x_i = 1 \right\}.$$

Let e^0 be the zero vector in $\mathbb{R}^n$. We obtain the n-dimensional standard simplex

$$\Delta_n \coloneqq \operatorname{Conv}(\{e^0, \ldots, e^n\}) = \left\{ x \in \mathbb{R}^n \colon x_i \geq 0, \sum_{i=1}^{n} x_i \leq 1 \right\}. \tag{1.1}$$

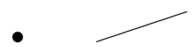

Figure 1.11: m-dimensional simplices: point $m = 0$ (left), and line segment $m = 1$ (right).

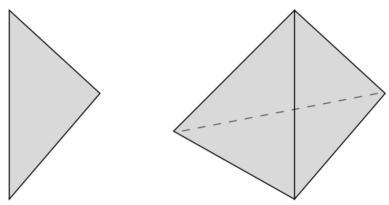

Figure 1.12: m-dimensional simplices: triangle $m = 2$ (left), and tetrahedron $m = 3$ (right).

Cones

A subset K of $\mathbb{R}^n$ is *conic* if K is nonempty and if $tx \in K$ for each $x \in K$ and $t \geq 0$. A convex conic set is called a *cone*. Figure 1.13 provides an illustration.

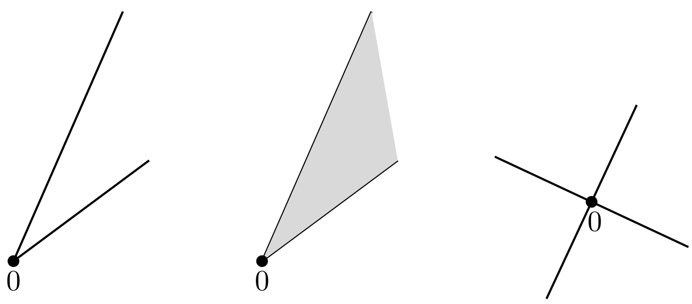

Figure 1.13: Conic set (left), cone (middle), and conic set (right). The dot is the origin 0.

Example 1.5. 1. The *nonnegative orthant*

$$\mathbb{R}^n_+ \coloneqq \{x \in \mathbb{R}^n \colon x \geq 0\} = \{x \in \mathbb{R}^n \colon x_i \geq 0, \quad i = 1, \ldots, n\}$$

is a cone.

2. The *Lorentz cone*

$$\mathbb{L}^n \coloneqq \left\{x \in \mathbb{R}^n \colon x_n \geq \sqrt{x_1^2 + \cdots + x_{n-1}^2}\right\} = \left\{(y,t) \in \mathbb{R}^{n-1} \times \mathbb{R} \colon t \geq \|y\|_2\right\}.$$

is a cone. Figure 1.14 depicts a graph of a Lorentz cone.

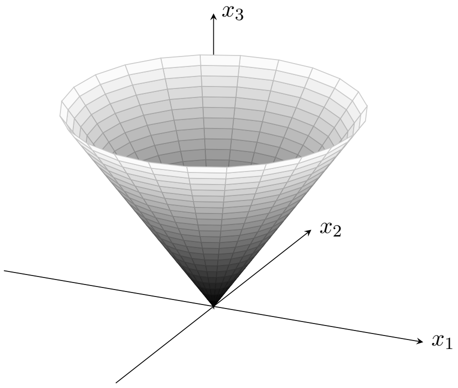

Figure 1.14: Graph of the boundary of the Lorentz cone $\mathbb{L}^3$.

3. The set of symmetric positive semidefinite matrices, called the semidefinite cone and denoted by $\mathbb{S}^n_+$, is a cone. We recall that a square matrix $A \in \mathbb{R}^{n \times n}$ is positive semidefinite if $d^T A d \geq 0$ for all $d \in \mathbb{R}^n$.
4. The solution set $\{x \in \mathbb{R}^n \colon a_\alpha^T x \leq 0, \quad \alpha \in \mathcal{A}\}$ of an arbitrary (finite or infinite) homogeneous system of nonstrict linear inequalities is a cone.

Cones can be characterized using the following fact.

Proposition 1.6. *A nonempty subset $K \subset \mathbb{R}^n$ is a cone if and only if*
1. *K is conic, that is, $tx \in K$ for each $x \in K$ and $t \geq 0$, and*
2. *K is closed with respect to addition, that is, $x + y \in K$ for each x, $y \in K$.*

Proof. Let K be a cone and let x, $y \in K$. Since K is convex, we have $(1/2)x+(1/2)y \in K$. Choosing $t = 2$, we find that $x + y \in K$.

Now let K be conic and closed with respect to addition. We have to show that K is convex. Let x, $y \in K$ and let $\lambda \in [0, 1]$. We have $\lambda x \in K$ and $(1 - \lambda)y \in K$, as K is conic. Since K is closed with respect to additions, we have $\lambda x + (1 - \lambda)y \in K$. □

Cones form an extremely important class of convex sets with properties related to those of general convex sets. For example,

- The intersection of an arbitrary familiy of cones is a cone. As a result, for every nonempty set $X \in \mathbb{R}^n$, the intersection of all cones containing X is a cone. This cone is the smallest cone containing X and is denoted by $\mathrm{Cone}(X)$ and called *conic hull* of $X \in \mathbb{R}^n$.
- A nonempty set $X \subset \mathbb{R}^n$ is a cone if and only if it equals the set of conic combinations of elements from X. A *conic combination* is a linear combination with nonnegative weights.
- The conic hull of a nonempty set $X \subset \mathbb{R}^n$ is the set of all conic combinations of element of X.

1.3 "Calculus" of Convex Sets

We provide a number of convexity-preserving operations.

Proposition 1.7. *The following operations preserve convexity of sets:*

1. Taking intersections of convex sets: *If $X_\alpha \subset \mathbb{R}^n$, $\alpha \in \mathcal{A}$, are convex, then $\cap_{\alpha\in\mathcal{A}} X_\alpha$ is convex.*
2. Taking direct products of convex sets: *If $X_\ell \subset \mathbb{R}^{n_\ell}$ for $\ell = 1, \ldots, L$ are convex, then their direct product*
$$X_1 \times \cdots \times X_L := \{(x^1, \ldots, x^L)\colon x^\ell \in X_\ell, \quad \ell = 1, \ldots, L\}$$
is convex.
3. Computing weighted sums of convex sets: *If $X_\ell \subset \mathbb{R}^{n_\ell}$ for $\ell = 1, \ldots, L$ are nonempty and convex, and $\lambda_\ell \in \mathbb{R}$, then*
$$\lambda_1 X_1 + \cdots + \lambda_L X_L := \{\lambda_1 x_1 + \cdots + \lambda_L x_L\colon x_\ell \in X_\ell, \quad \ell = 1, \ldots, L\}$$
is convex.
4. Taking the image of an affine mapping: *Let $X \subset \mathbb{R}^n$ be convex, let $A \in \mathbb{R}^{m\times n}$, and let $b \in \mathbb{R}^m$. Let us define the mapping $\mathcal{A}$ by $\mathcal{A}(x) := Ax + b$. Then the image of X under the mapping $\mathcal{A}$,*
$$\mathcal{A}(X) := \{y = Ax + b\colon x \in X\},$$
is convex.
5. Taking the inverse image under affine mapping: *Let $X \subset \mathbb{R}^n$ be convex, let $A \in \mathbb{R}^{n\times m}$, and let $b \in \mathbb{R}^n$. Let us define the mapping $\mathcal{A}$ by $\mathcal{A}(y) = Ay + b$. Then the inverse image of X under the mapping $\mathcal{A}$,*
$$\mathcal{A}^{-1}(X) := \{y \in \mathbb{R}^m\colon Ay + b \in X\},$$
is convex.

1.4 Closures and Interiors of Convex Sets

A set $X \subset \mathbb{R}^n$ is *closed* if for all sequences $(x^k) \subset X$ with $x^k \to x$ as $k \to \infty$, we have $x \in X$.

Example 1.8. 1. Unit norm balls, $X = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$, are closed. To establish this assertion, let $(x^k) \subset X$ be a sequence converging to x as $k \to \infty$. We have to show that $x \in X$. Using the triangle inequality, we can show that $\|x^k\| \to \|x\|$ as $k \to \infty$. Combined with $\|x^k\| \leq 1$ for all $k \in \mathbb{N}$, we obtain $\|x\| \leq 1$. Hence $x \in X$.

2. If $f : X \to \mathbb{R}$ is a continuous function on a nonempty, closed set $X \subset \mathbb{R}^n$, and $x^0 \in X$, then the level set

$$\{x \in X : f(x) \leq f(x^0)\}$$

is closed. Let us call this level set S. To show that S is closed, let (x^k) be a sequence contained in S with $x^k \to x$ as $k \to \infty$. Our task is to show that $x \in X$ and $f(x) \leq f(x^0)$. Since X is closed, we have $x \in X$. We have $f(x^k) \leq f(x^0)$ for all $k \in \mathbb{N}$. Since f is continuous and $x^k \to x$ as $k \to \infty$, we find that $f(x^0) \geq f(x^k) \to f(x)$ as $k \to \infty$. Hence $f(x) \leq f(x^0)$, yielding $x \in S$.

3. The set $X = \{x \in \mathbb{R}^n : a_\alpha^T x \leq b_\alpha,\, \alpha \in \mathcal{A}\}$ is closed. This is the solution set to an arbitrary system of nonstrict linear inequalities. We can demonstrate its closedness using Lemma 1.9.

A set $X \in \mathbb{R}^n$ is called *bounded* if there exists a number $r \in (0, \infty)$ with $\|x\|_2 \leq r$ for all $x \in X$. In other words, a set is bounded if and only if it is contained in a ball (about the origin) with finite radius. A set $X \in \mathbb{R}^n$ is *compact* if and only if it is bounded and closed.

Lemma 1.9. 1. *The intersection of an arbitrary family of closed sets is closed.*

2. *The union of a finite family of closed sets are closed.*

We can establish the fact that $X = \{x \in \mathbb{R}^n : a_\alpha^T x \leq b_\alpha,\, \alpha \in \mathcal{A}\}$ is closed using Lemma 1.9, because $X = \bigcap_{\alpha \in \mathcal{A}} \{x \in \mathbb{R}^n : a_\alpha^T x \leq b_\alpha\}$ and each of the sets $\{x \in \mathbb{R}^n : a_\alpha^T x \leq b_\alpha\}$ is closed.

A set $X \subset \mathbb{R}^n$ is *open* if for each $x \in X$, there exists $r > 0$ such that

$$\{y \in \mathbb{R}^n : \|y - x\|_2 < r\} \subset X.$$

In other words, a set $X \subset \mathbb{R}^n$ is open if for each point from X there exists a norm ball contained with positive radius in X. For example, the set $\{x \in \mathbb{R}^n : \|x\|_2 < 1\}$ is open.

A set $X \subset \mathbb{R}^n$ is open if and only if its complement $\mathbb{R}^n \setminus X$ is closed.

Lemma 1.10. *A set $X \subset \mathbb{R}^n$ is closed if and only if its complement $\mathbb{R}^n \setminus X$ is open.*

Using this lemma and Lemma 1.9, we obtain the following fact.

Lemma 1.11. 1. *The union of an arbitrary family of open sets is open.*

2. *The intersection of a finite family of open sets are open.*

The *closure* $\mathrm{cl}(X)$ of a set $X \subset \mathbb{R}^n$ is defined by

$$\mathrm{cl}(X) \coloneqq \{x \in \mathbb{R}^n \colon (x^k) \subset X,\, x^k \to x \text{ as } k \to \infty\}.$$

In other words, the closure of a set is the set of limit points of converging sequences of points in X. The closure of a set X is the smallest closed set containing X.

Example 1.12. 1. The closure of $\{x \in \mathbb{R}^n \colon \|x\|_2 < 1\}$ is $\{x \in \mathbb{R}^n \colon \|x\|_2 \leq 1\}$.
2. If $\{x \in \mathbb{R}^n \colon a_\alpha^T x < b_\alpha,\, \alpha \in \mathcal{A}\}$ is nonempty, then its closure is

$$\{x \in \mathbb{R}^n \colon a_\alpha^T x \leq b_\alpha,\, \alpha \in \mathcal{A}\}.$$

Let $X \subset \mathbb{R}^n$ be a set. A point $x \in X$ is an *interior point* of X if there exists $r > 0$ with

$$\{y \in \mathbb{R}^n \colon \|y - x\|_2 < r\} \subset X.$$

The set of all interior points of X is called *interior* of X and denoted by $\mathrm{int}(X)$. The interior of a set $X \subset \mathbb{R}^n$ is the largest open set contained in X. The interior of $\{x \in \mathbb{R}^n \colon \|x\|_2 \leq 1\}$ is $\{x \in \mathbb{R}^n \colon \|x\|_2 < 1\}$. The interior of the rational numbers is $\emptyset$.

Example 1.13. Let us consider the n-dimensional standard simplex defined in (1.1). The point $x = (1/(n+1), \ldots, 1/(n+1)) \in \mathbb{R}^n$ is an interior point of the standard simplex defined in (1.1).

The *boundary* ∂X of a set $X \subset \mathbb{R}^n$ is $\partial X \coloneqq \mathrm{cl}(X) \setminus \mathrm{int}(X)$. For example, the boundary of the sets $\{x \in \mathbb{R}^n \colon \|x\|_2 < 1\}$ and $\{x \in \mathbb{R}^n \colon \|x\|_2 \leq 1\}$ is the sphere $\{x \in \mathbb{R}^n \colon \|x\|_2 = 1\}$. For a set $X \subset \mathbb{R}^n$, we have

$$\mathrm{int}(X) \subset X \subset \mathrm{cl}(X).$$

Let us discuss some properties of closures and interiors of convex sets.

Proposition 1.14. 1. *If $X \subset \mathbb{R}^n$ is convex, then its interior and closure are convex.*
2. *If $X \subset \mathbb{R}^n$ is convex, $x \in \mathrm{int}(X)$, and $y \in \mathrm{cl}(X)$, then*

$$\lambda x + (1-\lambda)y \in \mathrm{int}(X) \quad \text{for all} \quad \lambda \in (0,1].$$

3. *If $X \subset \mathbb{R}^n$ is convex and its interior is nonempty, then*

$$\mathrm{cl}(X) = \mathrm{cl}(\mathrm{int}(X)). \tag{1.2}$$

Roughly speaking (1.2) says that each point from the closure of X can be "approximated" with points from the interior of X. The identity in (1.2) says that $\mathrm{int}(X)$ is dense in $\mathrm{cl}(X)$, provided that X is convex and has an interior point. If $X \subset \mathbb{R}^n$ is convex, closed, and its interior is nonempty, then (1.2) ensures

$$X = \mathrm{cl}(\mathrm{int}(X)).$$

The identity (1.2) does not hold if X as empty interior and X is nonempty. For example, if $X = \{x\}$ with $x \in \mathbb{R}^n$, then $\mathrm{int}(X) = \emptyset$. We obtain $\{x\} = X = \mathrm{cl}(X) \neq \mathrm{cl}(\emptyset) = \emptyset$.

Let us establish Proposition 1.14.

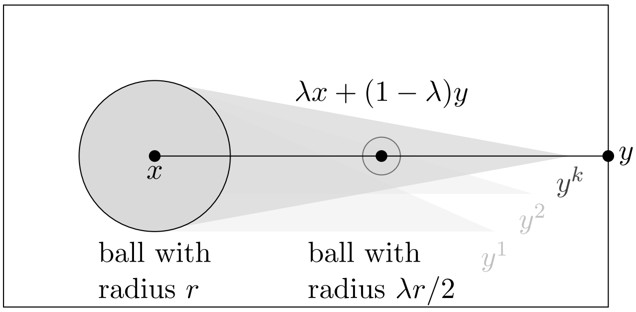

Figure 1.15: Rectangle set X, boundary point y, and interior point x. The ball about x is contained in the rectangle. As y^k approaches y, the ball about the convex combination with radius $\lambda r/2$ is eventually contained in the gray area.

Proof of Proposition 1.14. 1. Let $X \subset \mathbb{R}^n$ be convex. We show that its interior is convex. For $z \in \mathbb{R}^n$ and $r > 0$, we define the norm ball with radius r and center z by $B_r(z) := \{x \in \mathbb{R}^n : \|x - z\| < r\}$. Let x, $y \in \operatorname{int}(X)$ and let $\lambda \in [0, 1]$. Then there exists $r > 0$ such that the sets $B_r(x)$ and $B_r(y)$ are contained in X. Since X is convex, we have $\lambda X + (1-\lambda)X \subset X$. Here, we use the notion of weighted sums of sets defined in Proposition 1.7. Hence $\lambda B_r(x) + (1-\lambda)B_r(y) \subset X$. The set $\lambda B_r(x) + (1-\lambda)B_r(y)$ is equal to $B_r(\lambda x + (1-\lambda)y)$. Hence $\lambda x + (1-\lambda)y \in \operatorname{int}(X)$.

Next, we show that the closure of X is convex. Let x, $y \in \operatorname{cl}(X)$ and let $\lambda \in [0, 1]$. Since x, $y \in \operatorname{cl}(X)$, there exist sequences (x^k), $(y^k) \subset X$ with $x^k \to x$ and $y^k \to y$ as $k \to \infty$. The convexity of X ensures $\lambda x^k + (1-\lambda)y^k \in X$ for all $k \in \mathbb{N}$. Combined with

$$\lambda x^k + (1-\lambda)y^k \to \lambda x + (1-\lambda)y \quad \text{as} \quad k \to \infty,$$

we find that $\lambda x + (1-\lambda)y \in \operatorname{cl}(X)$.

2. Our proof approach is illustrated in Figure 1.15. Since $x \in \operatorname{int}(X)$, there exists $r > 0$ such that $B_r(x) \subset X$. Since $y \in \operatorname{cl}(X)$, there exists a sequence $(y^k) \subset X$ with $y^k \to y$ as $k \to \infty$. Fix $k \in \mathbb{N}$. We have $\lambda B_r(x) + (1-\lambda)y^k \subset X$, as $B_r(x)$ and X are convex. The set $\lambda B_r(x) + (1-\lambda)y^k$ is the norm ball about $\lambda x + (1-\lambda)y^k$ with radius λr. Hence $\lambda x + (1-\lambda)y^k \in \operatorname{int}(X)$. For all sufficiently large $k \in \mathbb{N}$, we have $\lambda B_{r/2}(x) + (1-\lambda)y \subset \lambda B_r(x) + (1-\lambda)y^k$. As the latter set is contained in the interior of X, so is $\lambda B_{r/2}(x) + (1-\lambda)y$. Hence $\lambda x + (1-\lambda)y \in \operatorname{int}(X)$.

3. We have $\operatorname{int}(X) \subset X$. Hence $\operatorname{cl}(\operatorname{int}(X)) \subset \operatorname{cl}(X)$. For the converse inclusion, we use the second part of the proposition. Let $x \in \operatorname{cl}(X)$. Then there exists a sequence $(x^k) \subset X$ with $x^k \to x$ as $k \to \infty$. Since the interior of X is nonempty, we can find $y \in \operatorname{int}(X)$. We observe that $(1/k)y + (1 - 1/k)x^k$ converges to x as $k \to \infty$. Using the second part of the proposition, we find that $(1/k)y + (1-1/k)x^k \in \operatorname{int}(X)$ for all $k \in \mathbb{N}$. Therefore, x is a limit point of a sequence of interior points.

□

If $X \subset \mathbb{R}^n$ is convex and its interior is nonempty, then

$$\operatorname{cl}(X) = \operatorname{cl}(\operatorname{int}(X)). \tag{1.3}$$

If $\operatorname{int}(X)$ is empty, this identity is generally wrong and as a result we may not approximate points in $\operatorname{cl}(X)$ by those in $\operatorname{int}(X)$. However, with a generalized notion of interior, we obtain an identity related to that in (1.3).

Definition 1.15. *Let $X \subset \mathbb{R}^n$ be a set. A point $x \in X$ is said to be a point in the relative interior of X if there exists $r > 0$ with*

$$\{y \in \mathbb{R}^n \colon \|y - x\|_2 < r\} \cap \operatorname{Aff}(X) \subset X.$$

The set of all such points is called the relative interior *and is denoted by* $\operatorname{rint}(X)$*. The* relative boundary *of X is the set* $\operatorname{cl}(X) \setminus \operatorname{rint}(X)$*.*

The relative interior of a singleton is the set itself. If $n > 1$, then the interior of a line segment between the points $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ is the empty set. In contrast, the relative interior of the line segment between x and y is the line segment without the points x and y. Figure 1.16 provides a two-dimensional illustration.

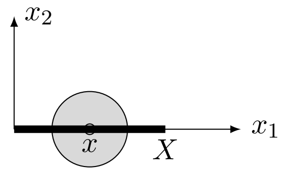

Figure 1.16: The set $X = \{x \in \mathbb{R}^2 \colon 0 \leq x_1 \leq 2,\, x_2 = 0\}$. The point $x = (1, 0)$ is in $\operatorname{rint}(X)$, but $\operatorname{int}(X) = \emptyset$.

Figure 1.17 provides a three-dimensional illustration.

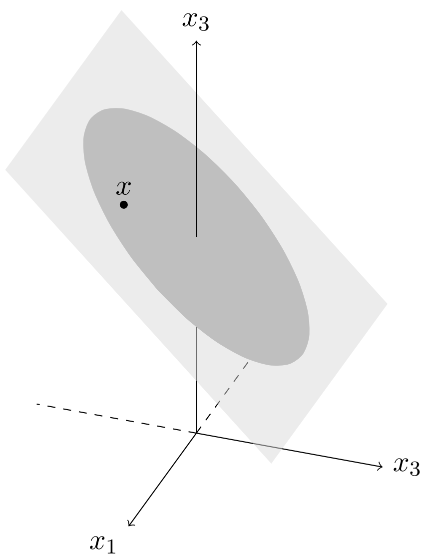

Figure 1.17: The dark gray area visualizes the convex set X. The light gray area visualizes the affine hull $\operatorname{Aff}(X)$ of the set X. The point x is in $\operatorname{rint}(X)$, but $\operatorname{int}(X) = \emptyset$.

Geometrically speaking, the relative interior $\operatorname{rint}(X)$ is the interior we obtain when viewing X as a subset of its affine hull $\operatorname{Aff}(X)$. The affine hull of a set is geometrically speaking nothing but $\mathbb{R}^m$, where m is the affine dimension of $\operatorname{Aff}(X)$. Let us provide details for the latter assertion. For a nonempty set $X \subset \mathbb{R}^n$, we have $X \subset \operatorname{Aff}(X)$ and $\operatorname{Aff}(X) = \{x \in \mathbb{R}^n \colon \exists z \in \mathbb{R}^m \text{ with } x = Bz + d\}$, where $d \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n\times m}$, and m is the affine dimension of $\operatorname{Aff}(X)$. So we have m degrees of freedom to generate points in $\operatorname{Aff}(X)$. Let us consider the set $\widehat{X} \coloneqq \{z \in \mathbb{R}^m \colon Bz + d \in X\}$. If $\bar{x} \in \operatorname{rint}(X)$, then $\bar{x}$ is an interior point of $\widehat{X}$. This fact is sometimes used in proofs of theorems.

For a set $X \subset \mathbb{R}^n$, we have

$$\operatorname{rint}(X) \subset X \subset \operatorname{cl}(X) \subset \operatorname{Aff}(X).$$

If $X \subset \mathbb{R}^n$ is a set and $\operatorname{Aff}(X) = \mathbb{R}^n$, then the relative interior of X equals the interior of X. In particular if the interior of X is nonempty, then we have $\operatorname{Aff}(X) = \mathbb{R}^n$. Indeed, if the interior of X is nonempty, then it contains a ball B with positive radius and we have $\mathbb{R}^n = \operatorname{Aff}(B) \subset \operatorname{Aff}(X)$.

The relative interior of a nonempty convex set is nonempty as we show next.

Theorem 1.16. 1. *If $X \subset \mathbb{R}^n$ is a nonempty convex set, then its relative interior is nonempty.*
2. *If $X \subset \mathbb{R}^n$ is convex, then $\operatorname{rint}(X)$ is convex.*

Proof. We have $\operatorname{rint}(X) = a + \operatorname{rint}(X - a)$ for each $a \in \mathbb{R}^n$. If $a \in X$, then we have $0 \in X - a$. Therefore, if X is nonempty, we can assume $0 \in X$. Otherwise, we shift X by some vector contained in X.

1. We use the above observation and assume $0 \in X$. Then we have $\operatorname{Aff}(X) = \operatorname{Lin}(X)$. If $X = \{0\}$, then we have $X = \operatorname{Lin}(X) = \{0\}$. Hence $\operatorname{rint}(X) = \{0\}$.

Now, let $X \neq \{0\}$. Since $\operatorname{Lin}(X)$ is a linear subspace, it has a basis $x^1, \dots, x^m \in X$ with m being the linear dimension of $\operatorname{Lin}(X)$. We have $m \geq 1$, as $X \neq \{0\}$. We consider the mapping $T : \mathbb{R}^m \to \operatorname{Lin}(X)$ defined by $T\lambda = \sum_{i=1}^m \lambda_i x^i$. This mapping takes weights $\lambda \in \mathbb{R}^m$ and outputs a linear combination of our basis vectors $x^1, \dots, x^m$. The mapping T is linear and continuous, and its inverse mapping T^{-1} is linear and continuous. If $U \subset \mathbb{R}^m$ is an open set, then the continuity of T ensures that $T(U) = \operatorname{Lin}(X) \cap V$ for some open set $V \subset \mathbb{R}^n$. Moreover, if $V \subset \mathbb{R}^n$ is open, then the continuity of T ensures that $T^{-1}(V) = \mathbb{R}^m \cap U$ for some open set $U \in \mathbb{R}^m$. We obtain

$$\operatorname{rint}(X) = T(\operatorname{int}(T^{-1}(X))). \tag{1.4}$$

The point $T^{-1}(x^i)$ equals the ith canonical unit vector e^i in $\mathbb{R}^m$ and $T^{-1}(0) = 0$. Proposition 1.7 and the convexity of X ensure that the inverse image $T^{-1}(X)$ is convex. Therefore, the m-dimensional standard simplex (see (1.1))

$$\Delta_m = \operatorname{Conv}(\{0, e^1, \dots, e^m\})$$

is contained in $T^{-1}(X)$. Example 1.13 tells us that Δ_m has an interior point $\bar{y}$. Hence $\bar{y} \in \operatorname{int}(S) \subset \operatorname{int}(T^{-1}(X))$ and we obtain $T(\bar{y}) \in T(\operatorname{int}(T^{-1}(X)))$. Combined with (1.4), we find that $\operatorname{rint}(X)$ is nonempty.

2. If X is empty, then $\operatorname{rint}(X)$ is empty and hence convex. Now let X be nonempty. We use the construction performed above and assume $0 \in X$. Proposition 1.7 and the convexity of X ensure that the inverse image $T^{-1}(X)$ is convex. Combined with Proposition 1.14, we find that $\operatorname{int}(T^{-1}(X))$ is convex. Applying Proposition 1.7 once more, we find that $T(\operatorname{int}(T^{-1}(X)))$ is convex. Combined with the identity in (1.4), we obtain the convexity of the relative interior of a convex set.

□

Using Theorem 1.16, we can extend the results in Proposition 1.14 to statements about the relative interior rather than the interior.

Proposition 1.17. *Let $X \subset \mathbb{R}^n$ be a nonempty convex set. Then*

$$\operatorname{cl}(X) = \operatorname{cl}(\operatorname{rint}(X)). \tag{1.5}$$

Moreover, if $x \in \operatorname{rint}(X)$, and $y \in \operatorname{cl}(X)$, then

$$\lambda x + (1-\lambda)y \in \operatorname{rint}(X) \quad \text{for all} \quad \lambda \in (0,1]. \tag{1.6}$$

Proof. Let us first establish (1.6). We use ideas similar to those used in the proof of Proposition 1.14. Let $x \in \operatorname{rint}(X)$. We have $\operatorname{Aff}(X) = a + L$ for some vector $a \in \mathbb{R}^n$ and a linear subspace L of $\mathbb{R}^n$. We have

$$X \subset \operatorname{Aff}(X) = x + L.$$

Let B be the closed unit ball in L, that is,

$$B \coloneqq \{h \in L \colon \|h\|_2 \leq 1\}.$$

Since $x \in \operatorname{rint}(X)$, there exists a positive radius $r > 0$ with

$$x + rB \subset X. \tag{1.7}$$

Now let $\lambda \in (0,1]$ and let $z = \lambda x + (1-\lambda)y$ with $y \in \operatorname{cl}(X)$. Since $y \in \operatorname{cl}(X)$, there exists a sequence $(y^k) \subset X$ with $y^k \to y$ as $k \to \infty$. Defining $z^k = \lambda x + (1-\lambda)y^k$, we have $z^k \to z$ as $k \to \infty$. Using (1.7) and the convexity of X, we find that the sets $Z_k = \{\lambda x' + (1-\lambda)y^k \colon x' \in X + rB\}$ are contained in X. Note that the set Z_k is the shifted ball $z^k + \lambda r B$. For every $r' < r$ and all $k \in \mathbb{N}$ such that z^k is sufficiently close to z, the ball $z^k + \lambda r B$ contains the ball $z + r'\lambda B$. Therefore, a neighborhood in $\operatorname{Aff}(X)$ of z belongs to M. Hence $z \in \operatorname{rint}(X)$, that is, (1.6) holds true.

Let us now establish (1.5). We use ideas similar to those in the proof of Proposition 1.14. We have $\operatorname{rint}(X) \subset X$. Hence $\operatorname{cl}(\operatorname{rint}(X)) \subset \operatorname{cl}(X)$. For the converse inclusion, we use (1.6). Let $x \in \operatorname{cl}(X)$. Then there exists a sequence $(x^k) \subset X$ with $x^k \to x$ as $k \to \infty$. Since the relative interior of X is nonempty according to Theorem 1.16, we can find $y \in \operatorname{rint}(X)$. We observe that $(1/k)y + (1-1/k)x^k$ converges to x as $k \to \infty$. Using (1.6), we find that $(1/k)y + (1-1/k)x^k \in \operatorname{rint}(X)$ for all $k \in \mathbb{N}$. Therefore, x is a limit point of a sequence of points from the relative interior. □

1.5 Caratheodory's theorem

We recall a few notions. A nonempty affine subspace M of $\mathbb{R}^n$ can be represented by $M = a + L$, where $a \in \mathbb{R}^n$ is a vector and $L \subset \mathbb{R}^n$ is a linear subspace. The linear subspace L has a basis. The number of elements in the basis is called linear dimension of L. The affine dimension of $M = a + L$ is the linear dimension of L. The affine dimension of a nonempty set $X \subset \mathbb{R}^n$ is the affine dimension of its affine hull $\operatorname{Aff}(X)$.

Let us illustrate these notions an specific examples.

Example 1.18.
- The affine dimension of $\{0\}$ is zero.
- The affine dimension of $\mathbb{R}^n$ is n.
- The affine dimension of the nonempty affine subspace $\{x \in \mathbb{R}^n \colon Ax = b\}$ is $n - \operatorname{Rank}(A)$. Here, $\operatorname{Rank}(A)$ is the rank of the matrix A.

With the notion of dimension of a set, we are ready to state Caratheodory's theorem.

Theorem 1.19 (Caratheodory's theorem)**.** *Let $X \subset \mathbb{R}^n$ be nonempty and let m be its affine dimension. Then each point $x \in \operatorname{Conv}(X)$ is a convex combination of at most $m+1$ points from X.*

Proof. Let $x \in \operatorname{Conv}(X)$. Then we have $x = \sum_{i=1}^p \lambda_i x^i$, for some $p \in \mathbb{N}$, $x^i \in X$ and positive weights $\lambda_i > 0$ with $\sum_{i=1}^p \lambda_i = 1$. (Per se we only know that x is a convex combination of elements of X. If some of the weights in the convex combination are zero, we remove the corresponding points from the convex combination. Therefore, we can assume $\lambda_i > 0$.) If the vectors $\lambda_i(x^i, 1) \in \mathbb{R}^{n+1}$, $i = 1, \ldots, p$, are linearly independent, then we must have $p \le m+1$. Let us provide further details for this claim. The maximum number affinely independent vectors in $\operatorname{Aff}(X)$ is $m+1$ and we have $x^i \in \operatorname{Aff}(X)$ for $i = 1, \ldots, p$. Moreover, $\lambda_i(x^i, 1) \in \mathbb{R}^{n+1}$, $i = 1, \ldots, p$, are linearly independent if and only if $(x^i, 1) \in \mathbb{R}^{n+1}$, $i = 1, \ldots, p$, are linearly independent if and only if the vectors $x^i \in \mathbb{R}^n$, $i = 1, \ldots, p$, are affinely independent if and only if the $p-1$ vectors $x^i - x^1 \in \mathbb{R}^n$, $i = 2, \ldots, p$, are linearly independent. Therefore, $p - 1 \le m$.

Now, let the vectors $y^i \coloneqq \lambda_i(x^i, 1) \in \mathbb{R}^{n+1}$, $i = 1, \ldots, p$, be linearly dependent. Hence, we can find $\mu \in \mathbb{R}^p \setminus \{0\}$ with $\sum_{i=1}^p \mu_i y^i = 0$. We obtain

$$\sum_{i=1}^p \mu_i \lambda_i = 0 \quad \text{and} \quad \sum_{i=1}^p \mu_i \lambda_i x^i = 0.$$

Since $\lambda_i > 0$ for $i = 1, \ldots, p$ and $\mu \neq 0$, at least one component of μ must be negative. We choose $j \in \mathbb{N}$ with $\mu_j = \min_i \mu_i$ and define $\lambda_i^* = (1 - \mu_i/\mu_j)\lambda_i$. We have $\mu_i/\mu_j \le 1$ and hence $1 \ge \lambda_i^* \ge 0$. Moreover, $\lambda_j^* = 0$. We compute

$$\sum_{i \neq j} \lambda_i^* = \sum_{i=1}^p \lambda_i^* = \sum_{i=1}^p \lambda_i - (1/\mu_j) \sum_{i=1}^p \mu_i \lambda_i = 1 - 0,$$

$$\sum_{i \neq j} \lambda_i^* x^i = \sum_{i=1}^p \lambda_i^* x^i = \sum_{i=1}^p \lambda_i x^i - (1/\mu_j) \sum_{i=1}^p \mu_i \lambda_i x^i = x - 0.$$

We have shown that x can be written as a convex combinations of $p-1$ vectors. Repeating our computations until $p \leq m+1$, we obtain the assertion. $\square$

Figure 1.18 provides an example showing that a convex combination with fewer addends than $m+1$ may not be possible.

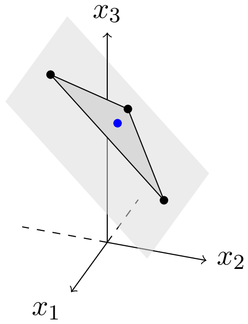

Figure 1.18: The blue point is the barycenter of the three triangle's vertices. The affine dimension of the affine hull of the triangle's vertices is two.

Example 1.20 (Application of Caratheodory's theorem)**.** A supermarket sells 99 different herbal teas. Each of them is a certain blend of 26 herbs. In spite of such a variety of blends, a customer, John, is not satisfied with any of them. The only herbal tea he likes is their mixture in the proportion

$$1\colon 2\colon 3\colon \cdots\colon 98\colon 99.$$

Once it occurred to John that in order to prepare his favorite tea, he does not need to buy all 99 teas. What is the number of herbal teas John must buy in order to great his favorite tee? The answer is just 26.

Let us represent a blend by a unit weight portion, say, 1 gram. Such a portion can be identified with a 26-dimensional vector $x = (x_1, \ldots, x_{26})$ with nonnegative entries summing to 1. Each x_i is the weight (in grams) of herb #i in the portion. We have

$$x \in \mathbb{R}^{26}_{+} \quad \text{and} \quad \sum_{i=1}^{26} x_i = 1.$$

Let $x^1, \ldots, x^{99}$ be the 99 herbal teas, each of them is a 26-dimensional vector. To create John's favorite tea $\bar{x}$, we mix the herbal teas. For each herbal tea x^i, we take λ_i grams of it and mix them together. We obtain

$$\bar{x} = \sum_{i=1}^{99} \lambda_i x^i.$$

We find that $\sum_{i=1}^{99} \lambda_i = 1$. Hence John's tea $\bar{x}$ can be obtained by mixing the herbal teas $x^1, \ldots, x^{99}$ if and only if $\bar{x} \in \operatorname{Conv}(\{x^1, \ldots, x^{99}\})$. Caratheodory's theorem tells us that John's tea $\bar{x}$ can be obtained by some mixture of at most $m+1$ of the herbal teas, where m is the affine dimension of the affine span of $x^1, \dots, x^{99}$. The affine span of the herbal teas belongs to the 25-dimensional affine plane

$$\left\{ x \in \mathbb{R}^{26} \colon \sum_{i=1}^{26} x_i = 1 \right\}$$

Hence $m \leq 25$.

We continue with the conic version of Caratheodory's theorem.

Theorem 1.21 (Caratheodory's theorem (conic version))**.** *Let $X \subset \mathbb{R}^n$ be nonempty. Then each point $x \in \operatorname{Cone}(X)$ is a conic combination of at most n points from X.*

Proof. Let $x \in \operatorname{Cone}(X)$. Since $x \in \operatorname{Cone}(X)$, we have $x = \sum_{i=1}^m \lambda_i x^i$ for some $\lambda_i > 0$ and $x^i \in X$. If $m \leq n$, then the statement is true. Now let $m \geq n+1$. Then the vectors $\lambda_1 x^1, \dots, \lambda_m x^m$ are linearly dependent, as the vectors $x^1, \dots, x^m$ are linearly dependent and $\lambda_i > 0$. Hence there exists $\mu \in \mathbb{R}^m$ with $\mu \neq 0$ such that $\sum_{i=1}^m \lambda_i \mu_i x^i = 0$. Since $\mu \neq 0$, at least one component of μ is nonzero. Suppose that μ has a negative entry. (If this is not the case, then we can multiply $\sum_{i=1}^m \lambda_i \mu_i x^i = 0$ by minus one, to obtain that at least on component of $-\mu$ is negative.) We choose $j \in \mathbb{N}$ with $\mu_j = \min_i \mu_i$ and define $\lambda_i^* = (1 - \mu_i/\mu_j)\lambda_i$. We have $\mu_i/\mu_j \leq 1$ and hence $\lambda_i^* \geq 0$ and $\lambda_j^* = 0$. We obtain $x = \sum_{i=1}^m \lambda_i x^i = \sum_{i=1}^m \lambda_i x^i - (1/\mu_j) \sum_{i=1}^m \mu_i \lambda_i x^i = \sum_{i=1}^m \lambda_i (1 - \mu_i/\mu_j) x^i = \sum_{i \neq j} \lambda_i^* x^i$. Hence, we have written x as a conic combination of $m-1$ points. Repeating this argument until $m \leq n$, we obtain the assertion. □

The example depicted in Figure 1.19 shows that a conic combination with fewer addends than n may not be possible.

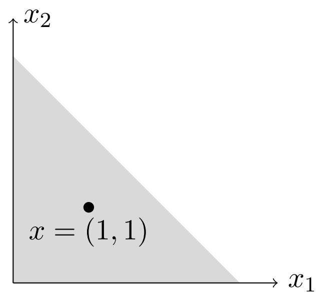

Figure 1.19: The set $X = \{(1,0),(0,1)\} \subset \mathbb{R}^2$ and its conic hull $\operatorname{Cone}(X) = \mathbb{R}^2_+$. The point $x = (1,1)$ is a conic combination of $(1,0)$ and $(0,1)$.

1.6 Radon's theorem

We recall that two sets are called disjoint if their intersection is empty.

Theorem 1.22 (Radon's theorem)**.** *Every collection of affinely dependent vectors in $\mathbb{R}^n$ can be split into two nonempty, disjoint sets such that their convex hulls have at least one point in common.*

Proof. Let $x^1, \ldots, x^m$ be a collection of m affinely dependent points. Then there exist weights $\lambda_i \in \mathbb{R}$ for $i = 1, \ldots, m$ with $\lambda \neq 0$,

$$\sum_{i=1}^{m} \lambda_i x^i = 0 \quad \text{and} \quad \sum_{i=1}^{m} \lambda_i = 0.$$

Let us define the index sets

$$I \coloneqq \{i \in \{1, \ldots, m\} \colon \lambda_i > 0\} \quad \text{and} \quad J \coloneqq \{i \in \{1, \ldots, m\} \colon \lambda_i \leq 0\}.$$

The set J is the complement of I. Since $\lambda \neq 0$ and the components of λ sum up to zero, the sets I and J are nonempty. Our definition of the sets ensures that they are disjoint. We obtain

$$\sum_{i \in I} \lambda_i x^i = -\sum_{j \in J} \lambda_j x^j = \sum_{j \in J} \underbrace{(-\lambda_j)}_{\geq 0} x^j$$

and

$$\sum_{i \in I} \lambda_i = -\sum_{j \in J} \lambda_j = \sum_{j \in J} (-\lambda_j) > 0.$$

We define $\sigma \coloneqq \sum_{j \in J} (-\lambda_j)$. We obtain

$$\operatorname{Conv}(\{x^i \colon i \in I\}) \ni \sum_{i \in I} (\lambda_i/\sigma) x^i = \sum_{j \in J} (-\lambda_j/\sigma) x^j \in \operatorname{Conv}(\{x^j \colon j \in J\}).$$

□

The following consequence of Radon's theorem (Theorem 1.22) is also known as Radon's theorem.

Corollary 1.23 (Radon's theorem)**.** *Every collection of m vectors in $\mathbb{R}^n$ with $m \geq n+2$ can be split into two nonempty, disjoint sets such that their convex hulls have at least one point in common.*

We establish Caratheodory's theorem (Theorem 1.19) using Theorem 1.22. Let $x \in \operatorname{Conv}(X)$. Then $x = \sum_{i=1}^{p} \lambda_i x^i$, for some $p \in \mathbb{N}$, $x^i \in X$ and positive weights $\lambda_i > 0$ with $\sum_{i=1}^{p} \lambda_i = 1$. If $p \leq m+1$, then the assertion is true. If $p > m+1$, then Radon's theorem ensures the existence of two nonempty, disjoint sets I, $J \subset \{1, \ldots, p\}$ and $\mu_i > 0$, $i = 1, \ldots, p$, $\sum_{i \in I} \mu_i = 1$, $\sum_{j \in J} \mu_j = 1$, such that

$$\sum_{i \in I} \mu_i x^i = \sum_{j \in J} \mu_j x^j \in \operatorname{Conv}(\{x^j \colon j \in J\}) \cap \operatorname{Conv}(\{x^i \colon i \in I\}).$$

We define $t \coloneqq \min_{i\in I}(\lambda_i/\mu_i)$. We have $\lambda_i \geq t\mu_i$ for all $i \in I$, $\lambda_k = t\mu_k$ for some $k \in I$, $t > 0$, and

$$x = x - t\sum_{i\in I}\mu_i x^i + t\sum_{j\in J}\mu_j x^j = \sum_{i\in I\setminus\{k\}}(\lambda_i - t\mu_i)x^i + \sum_{j\in J}(\lambda_j + t\mu_j)x^j$$

and

$$\sum_{i\in I}(\lambda_i - t\mu_i) + \sum_{j\in J}(\lambda_j + t\mu_j) = 1.$$

Repeating this argument until $p \leq m+1$, we obtain the assertion.

1.7 Helly's Theorem

We establish Helly's theorem using Radon's theorem.

Theorem 1.24 (Helly's theorem)**.** *Let $X_1, \dots, X_M$ be convex sets in $\mathbb{R}^n$ with $M \geq n+1$. Suppose that every $n+1$ sets from the family $X_1, \dots, X_M$ have a point in common. Then all sets have a point in common.*

Proof. We establish the assertion using induction and Radon's theorem.

If $M = n+1$, then the sets $X_1, \dots, X_M$ have a point in common by assumption.

Now we perform the induction step. Suppose that our assertion holds true for every M members of the family and let us show that the statement is true for the family $X_1, \dots, X_{M+1}$.

By induction hypotheses, every one of the $M+1$ sets

$$B_\ell \coloneqq X_1 \cap X_2 \cap \cdots \cap X_{\ell-1} \cap X_{\ell+1} \cap \cdots \cap X_{M+1} = \bigcap_{i\neq\ell} X_i$$

is nonempty. Let us choose for every $\ell = 1, \dots, M+1$, a point $x^\ell \in B_\ell$.

Radon's theorem ensures that the collection $x^1, \dots, x^{M+1}$ can be split into two nonempty, distinct sets and their convex hulls intersect. We can assume without loss of generality that these two sets are given by $\{x^1, \dots, x^{J-1}\}$ and $\{x^J, \dots x^{M+1}\}$ with $J \in \{1, \dots, M+1\}$ and that

$$z \in \operatorname{Conv}(\{x^1, \dots, x^{J-1}\}) \cap \operatorname{Conv}(\{x^J, \dots x^{M+1}\}).$$

We claim that $z \in X_\ell$ for $\ell = 1, \dots, M+1$. The points $x^J, \dots x^{M+1}$ are contained in the sets X_ℓ for $\ell = 1, \dots, J-1$. Combined with the fact that the sets X_ℓ are convex, we obtain

$$z \in \operatorname{Conv}(\{x^J, \dots x^{M+1}\}) \subset X_\ell, \quad \ell = 1, \dots, J-1.$$

Moreover, the points $x^1, \dots, x^{J-1}$ are contained in the sets X_ℓ for $\ell = J, \dots, M+1$. Hence

$$z \in \operatorname{Conv}(\{x^1, \dots, x^{J-1}\}) \subset X_\ell, \quad \ell = J, \dots, M+1.$$

Putting together the pieces, we have shown that $z \in X_\ell$ for $\ell = 1, \dots, M+1$. □

Example 1.25. We consider the Tschebyshev problem

$$\min_{x\in\mathbb{R}^n} \max_{1\le i\le M} |a_i^T x - b_i| \tag{1.8}$$

where $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ with and $M \geq n+1$. Let v^* be the optimal value of (1.8). We show that there is an index set $I \subset \{1,\dots,M\}$ with at most $n+1$ indices such that the optimal value of the relaxed problem

$$\inf_{x\in\mathbb{R}^n} \max_{i\in I} |a_i^T x - b_i| \tag{1.9}$$

equals the optimal value v^* of (1.8).

We consider the sets $X_i = \{x \in \mathbb{R}^n \colon |a_i^T x - b_i| < v^*\}$. Since v^* is the optimal value of (1.8), the sets $X_1,\dots,X_M$ have no point in common. (If they would have a point in common, then v^* would not be the optimal value of (1.8).) Helly's theorem ensures that there exists an index set $I \subset \{1,\dots,M\}$ with at most $n+1$ elements such that X_i, $i \in I$, have no point in common. In other words, there is no point $\bar{x} \in \mathbb{R}^n$ that satisfies all inequalities $|a_i^T \bar{x} - b_i| < v^*$ for all $i \in I$. We obtain that the optimal value of (1.9) is $\geq v^*$. Since $I \subset \{1,\dots,M\}$, these optimal values must be equal.

Helly's theorem can be refined.

Theorem 1.26. *Let $X_1,\dots,X_M$ be a family of convex sets in $\mathbb{R}^n$. Suppose that their union $X_1 \cup X_2 \cup \dots \cup X_M$ is contained in an affine subspace of affine dimension m. If every $m+1$ sets from the family $X_1,\dots,X_M$ have a point in common, then the sets $X_1,\dots,X_M$ have a point in common.*

Proof. The sets X_j can be viewed as subsets of $\mathbb{R}^m$. Therefore, Helly's theorem implies the assertion. □

We consider some applications of Helly's theorem.

Example 1.27 (Polynomial approximation)**.** We are given a function $f : X \to \mathbb{R}$ defined on a set $X \subset \mathbb{R}$ with 7,000,000 points. For each subset of X with 7 points, the function f can be approximated with a polynomial of degree 5 with accuracy 0.001. To approximate the function f on X, we want to use a spline of degree 5. A spline of degree 5 is a piecewise polynomial with pieces of degree 5. How many pieces should we take to approximate f by a spline with accuracy 0.001 on X? The answer is one is sufficient.

For $x \in X$, let A_x be the set of coefficients of all polynomials of degree 5 which approximate $f(x)$ within accuracy 0.001, that is,

$$A_x = \left\{ p = (p_0,\dots,p_5) \in \mathbb{R}^6 \colon \left| f(x) - \sum_{i=0}^{5} p_i x^i \right| \leq 0.001 \right\}.$$

The set A_x is polyhedral and hence convex. By assumption, every $6+1=7$ sets from the family $(A_x)_{x\in X}$ have a point in common. Helly's theorem implies that all sets $(A_x)_{x\in X}$ have a point in common. Hence there exists on polynomial of degree 5 which approximates f within accuracy 0.001 at every point in X.

Example 1.28. Our task is the design a factory described by the following linear programming model:

$$\begin{aligned} Ax &\geq d \quad && d_1, \ldots, d_{1000}, \text{demands} \\ Bx &\leq f \quad && f_1 \geq 0, \ldots, f_{10} \geq 0, \text{amount of resources} \\ Cx &\leq c, \quad && \text{other constraints} \end{aligned} \tag{1.10}$$

The problem data A, B, C, c is given. Our task is to order resources $f_i \geq 0$, $i = 1, \ldots, 10$, in such a way that the factory will be capable of satisfying all demand scenarios d from a given finite set D, that is, the system (1.10) should have a feasible point for every $d \in D$. The ith resource costs $a_i f_i$ with $a_i > 0$. It is known to us that in order to satisfy a scenario $d \in D$, it suffices to invest \$1 in the resources $f_1, \ldots, f_{10}$. How large should the investment in resources be in case that D contains

- ten scenarios?
- 2004 scenarios?

If D consists of 10 possible scenarios, then an investment of \$10 is sufficient. If D consists of 2004 possible scenarios, then an investment of \$11 is sufficient. Let us establish this statement. For $d \in D$, let F_d be the set of all nonnegative $f \in \mathbb{R}^{10}$ with cost at most \$11 and result in a solvable system

$$Ax \geq d, \quad Bx \leq f, \quad Cx \leq c \tag{1.11}$$

with x as the unknown. The set F_d is convex as a result of the representation

$$F_d = \{f \in \mathbb{R}^{10}_+ \colon \exists x \quad \text{solving} \quad (1.11)\}$$

and Theorem 1.30, for example. Every 11 sets of this type have a point in common: Given 11 scenarios $d^1, \ldots d^{11} \in D$, we can meet demand scenario d^i by investing \$1 in a properly selected vector of resources $f^i \in \mathbb{R}^{10}_+$. Therefore, we can meet the demand of the 11 scenarios $d^1, \ldots d^{11}$ by a single vector of resources $f^1 + \cdots + f^{11}$ at the cost of \$11. Therefore, this vector belongs to every one of the sets $F_{d^1}, \ldots F_{d^{11}}$. Since very 11 of the 2004 convex sets $F_d \subset \mathbb{R}^{10}$, $d \in D$, have a point in common, Helly's theorem ensures that all of these sets have a point in common.

Helly's theorem has important applications in structural design, for example, with interesting implications.

1.8 Polyhedral Representations

We recall that a set X in $\mathbb{R}^n$ is called polyhedral if it is the solution set to a finite system of nonstrict linear inequalities, that is, $X \subset \mathbb{R}^n$ is polyhedral if it takes the form

$$X = \{x \in \mathbb{R}^n \colon Ax \leq b\} = \{x \in \mathbb{R}^n \colon a_i^T x \leq b_i, \quad i = 1, \ldots, m\}.$$

We say that a set $X \subset \mathbb{R}^n$ is *polyhedrally representable* if it admits a representation of the form

$$X = \{x \in \mathbb{R}^n \colon \exists w \in \mathbb{R}^m \quad \text{with} \quad Px + Qw \leq r\}, \tag{1.12}$$

where $P \in \mathbb{R}^{s\times n}$, $Q \in \mathbb{R}^{s\times m}$, and $r \in \mathbb{R}^s$. A respresentation of a set $X \subset \mathbb{R}^n$ of the form in (1.12) is called *polyhedral representation* of X, and the variables w in (1.12) are called *slack variables*. The set (1.12) is the projection onto the space of x-variables of the polyhedral set

$$X^+ = \{(x, w) \in \mathbb{R}^{n+p} \colon Px + Qw \le r\}.$$

Figure 1.20 depicts an illustration.

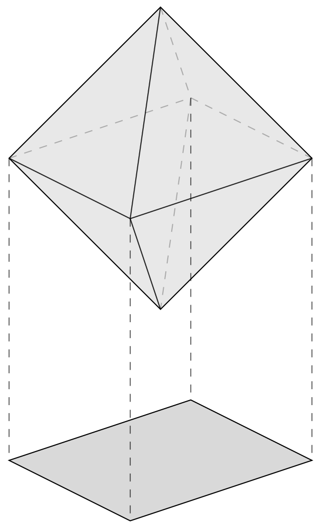

Figure 1.20: A polyhedron (top) and its projection (bottom).

Example 1.29. 1. Every polyhedral set $X = \{x \in \mathbb{R}^n \colon Ax \le b\}$ is polyhedrally representable. In (1.12), we can choose $P = A$ and $Q = 0$.

2. The set $X = \{x \in \mathbb{R}^n \colon \sum_{i=1}^n |x_i| \le 1\}$ is polyhedrally representable. We have

$$X = \left\{x \in \mathbb{R}^n \colon \exists w \in \mathbb{R}^n \text{ with } -w_i \le x_i \le w_i, 1 \le i \le n, \sum_{i=1}^n u_i \le 1\right\}.$$

To establish this identity it is helpful to use the fact that $-w_i \le x_i \le w_i$ if and only if $|x_i| \le w_i$.

3. The conic hull of the vectors $a_1, \dots, a_m \in \mathbb{R}^n$ can be represented as a polyhedron, because

$$\begin{aligned}\operatorname{Cone}(\{a_1, \dots, a_m\}) &= \left\{x \in \mathbb{R}^n \colon \exists \lambda \in \mathbb{R}^m_+ \quad \text{with} \quad x = \sum_{i=1}^n \lambda_i a_i\right\} \\ &= \left\{x \in \mathbb{R}^n \colon \exists \lambda \in \mathbb{R}^m \text{ with } \begin{cases} -\lambda & \le 0 \\ x - \sum_{i=1}^n \lambda_i a_i & \le 0 \\ -x + \sum_{i=1}^n \lambda_i a_i & \le 0 \end{cases}\right\}\end{aligned}$$

It turns out that every polyhedrally representable set is polyhedral; a surprising and deep fact.

Theorem 1.30. *Every polyhedrally representable set is polyhedral.*

Proof. The proof is constructive and uses the Fourier–Motzkin elimination scheme. Let us consider the set

$$X = \{x \in \mathbb{R}^n \colon \exists w \in \mathbb{R} \quad \text{with} \quad Px + wq \le r\},$$

where $P \in \mathbb{R}^{s\times n}$, $q \in \mathbb{R}^s$, and $r \in \mathbb{R}^s$. This set has only one slack variable.

Let us show that this sets can be represented as a polyhedral set. We define the index sets $I_- \coloneqq \{i \colon q_i < 0\}$, $I_0 \coloneqq \{i \colon q_i = 0\}$, and $I_+ \coloneqq \{i \colon q_i > 0\}$. Note that some of these three index sets may be empty. Let us denote the ith row of P by p_i^T. We have

$$\begin{aligned}
x \in X &\Leftrightarrow \exists w \in \mathbb{R} \colon \quad Px - r \le -wq \\
&\Leftrightarrow \exists w \in \mathbb{R} \colon p_i^T x - r_i \le -wq_i, \quad i \in \{1, \dots, s\} \\
&\Leftrightarrow \exists w \in \mathbb{R} \colon \begin{cases} p_i^T x - r_i \le -wq_i, & i \in I_-, \\ p_j^T x - r_j \le -wq_j, & j \in I_+, \\ p_k^T x - r_k \le 0, & k \in I_0 \end{cases} \\
&\Leftrightarrow \exists w \in \mathbb{R} \colon \begin{cases} (r_i - p_i^T x)/q_i \le w, & i \in I_-, \\ (r_j - p_j^T x)/q_j \ge w, & j \in I_+, \\ p_k^T x - r_k \le 0, & k \in I_0 \end{cases} \\
&\Leftrightarrow \begin{cases} (r_i - p_i^T x)/q_i \le (r_j - p_j^T x)/q_j, & i \in I_-, j \in I_+, \\ p_k^T x - r_k \le 0, & k \in I_0, \end{cases}
\end{aligned}$$

Let us define $a_i = p_i/q_i$ for $i \in I_+ \cup I_-$ and $b_i = r_i/q_i$ for $i \in I_+ \cup I_-$. We obtain

$$X = \{x \in \mathbb{R}^n \colon (a_j - a_i)^T x \le b_j - b_i, \;\; i \in I_-, j \in I_+, \quad p_k^T x \le r_k, \;\; k \in I_0\}.$$

This set is polyhedral.

So far we have considered a set with only one slack variable. If X is as in (1.12), then we only need to apply the above elimination scheme m times. □

The fact that every polyhedrally representable set is polyhedral has many applications. We provide here one. We consider the linear program (LP)

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax \le b, \tag{1.13}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ is a matrix, and $b \in \mathbb{R}^m$. Let us consider the set

$$X = \{\tau \in \mathbb{R} \colon \exists x \in \mathbb{R}^n \quad \text{with} \quad Ax \le b, \quad c^T x = \tau\}.$$

Converting $c^T x = \tau$ into two linear inequalities, we obtain

$$X = \{\tau \in \mathbb{R} \colon \exists x \in \mathbb{R}^n \quad \text{with} \quad Ax \leq b, \quad c^T x \leq \tau \quad -c^T x \leq -\tau\}.$$

According to Theorem 1.30, this set is polyhedral. In particular, X can be represented by a finite system of linear inequalities in the variable τ. By definition of the set X, it is the set of values the objective function $\tau = c^T x$ for feasible $x \in \mathbb{R}^n$ (those x that satisfy $Ax \leq b$). We find that if the LP (1.13) has a feasible point and its objective function is bounded from below on the feasible set, then X has a smallest element. We deduce the existence of a solution to the LP (1.13), provided that it has a feasible point and its objective function is bounded from below on the feasible set.

1.9 Separation Theorems

Separation theorems make statements about whether or not convex sets can be separated by a hyperplane.

Definition 1.31. *Let $a \in \mathbb{R}^n \setminus \{0\}$ and let $b \in \mathbb{R}$. The set*

$$H(a, b) \coloneqq \{x \in \mathbb{R}^n \colon a^T x = b\}.$$

is called a hyperplane.

Hyperplanes are $(n-1)$-dimensional affine subspaces in $\mathbb{R}^n$. Figures 1.21 and 1.22 provides illustrations.

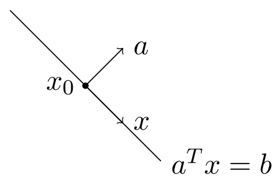

Figure 1.21: A hyperplane.

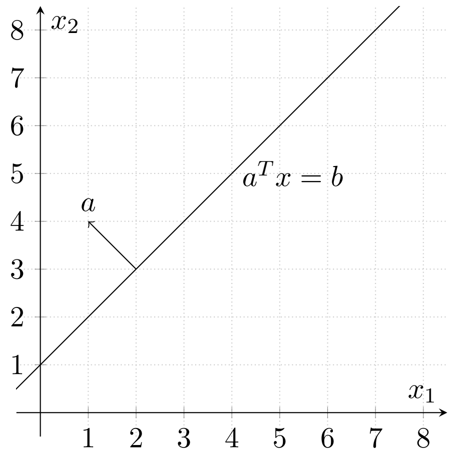

Figure 1.22: A hyperplane with $a = (-1, 1)$.

For $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$, we define the closed *halfspaces*

$$H_-(a,b) := \{x \in \mathbb{R}^n \colon a^T x \leq b\},$$
$$H_+(a,b) := \{x \in \mathbb{R}^n \colon a^T x \geq b\}.$$

If $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$, then $\mathbb{R}^n = H_-(a,b) \cup H_+(a,b)$. Figure 1.23 provides an illustration.

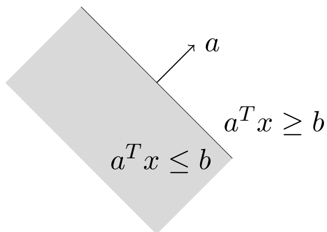

Figure 1.23: Halfspaces.

We formalize the concept of separation by hyperplanes.

Definition 1.32. *Let T and S be two nonempty sets in $\mathbb{R}^n$. Let $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$.*

1. *The hyperplane $H(a,b)$* separates *S and T if*

$$S \subset H_-(a,b), \quad T \subset H_+(a,b), \quad \text{and} \quad S \cup T \not\subset H(a,b).$$

(The last condition says that at least one of the sets S and T is not contained in $H(a,b)$.)

2. *We say that a linear form $a^T x$ with $a \neq 0$* separates *S and T if for some $b \in \mathbb{R}$, the hyperplane $H(a,b)$ separates S and T.*

3. *The hyperplane* $H(a, b)$ *separates* S *and* T strongly *if there exist* b' *and* b'' *with* $b' < b < b''$ *such that*

$$S \subset H_-(a, b') \quad \textit{and} \quad T \subset H_+(a, b'').$$

4. *We say that* S *and* T *can be* separated strongly *if there exists a hyperplane separating* S *and* T *strongly.*

Figure 1.24 provides an illustration.

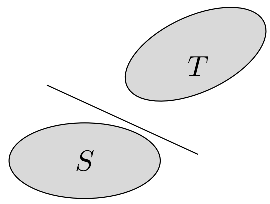

Figure 1.24: The hyperplane $2x_1 + 3x_2 = 6$ separates S and T strongly.

An important characterization of a separating linear form is given as follows. Let S and T be nonempty convex sets in $\mathbb{R}^n$ and let $a \in \mathbb{R}^n \setminus \{0\}$. Then the linear form $a^T x$ separates S and T if and only if

$$\sup_{x \in S} a^T x \leq \inf_{y \in T} a^T y \quad \text{and} \quad \inf_{x \in S} a^T x < \sup_{y \in T} a^T y. \tag{1.14}$$

Moreover, the linear form $a^T x$ strongly separates S and T if and only if

$$\sup_{x \in S} a^T x < \inf_{y \in T} a^T y. \tag{1.15}$$

If S and T are nonempty, and $a \in \mathbb{R}^n \setminus \{0\}$ satisfies (1.14), then for each $b \in \mathbb{R}$ with

$$\sup_{x \in S} a^T x \leq b \leq \inf_{y \in T} a^T y, \tag{1.16}$$

the hyperplane $H(a, b)$ separates the sets S and T.

Example 1.33. 1. The linear form x_1 on $\mathbb{R}^2$ separates the sets

$$S = \{x \in \mathbb{R}^2 \colon x_1 \leq 0,\ x_2 \leq 0\},$$
$$T = \{x \in \mathbb{R}^2 \colon x_1 \geq 0,\ x_2 \geq 0\}.$$

2. The linear form x_1 on $\mathbb{R}^2$ separates the sets

$$S = \{x \in \mathbb{R}^2 \colon x_1 \leq 0,\ x_2 \leq 0\},$$
$$T = \{x \in \mathbb{R}^2 \colon x_1 + x_2 \geq 0,\ x_2 \leq 0\}.$$

3. The linear form x_1 on $\mathbb{R}^2$ does not separate the sets
$$S = \{x \in \mathbb{R}^2 \colon x_1 = 0,\ 1 \le x_2 \le 2\},$$
$$T = \{x \in \mathbb{R}^2 \colon x_1 = 0,\ -2 \le x_2 \le -1\}.$$
4. The linear form x_1 on $\mathbb{R}^2$ separates the sets
$$S = \{x \in \mathbb{R}^2 \colon x_1 = 0,\ 0 \le x_2 \le 2\},$$
$$T = \{x \in \mathbb{R}^2 \colon 0 \le x_1 \le 1,\ -2 \le x_2 \le 1\}.$$

1.9.1 Separation Theorem

Theorem 1.34 provides a separation theorem.

Theorem 1.34 (separating hyperplane theorem)**.** *Two nonempty convex sets in $\mathbb{R}^n$ can be separated by a hyperplane if and only if their relative interiors do not intersect.*

We prepare our proof of Theorem 1.34. The following lemma provides conditions necessary and sufficient for a linear form over a convex set be constant. We will use this lemma to show that if two nonempty convex sets can be separated by a hyperplane, then their relative interiors have no point in common.

Lemma 1.35. *Let $X \subset \mathbb{R}^n$ be a convex set, $\bar{x} \in \operatorname{rint}(X)$, and let $a \in \mathbb{R}^n$. Then $a^T\bar{x} = \max_{x\in X} a^Tx$ if and only if the linear form a^Tx is constant on X.*

Proof. If the linear form a^Tx is constant on X, then any point $x \in X$ is a maximizer of a^Tx over X. We have $a^Tx = a^T\bar{x}$ for all $x \in X$.

Now, let $\bar{x} \in \operatorname{rint}(X)$ be a maximizer of a^Tx over X and let $y \in X$. We should prove that $a^T\bar{x} = a^Ty$. If $y = \bar{x}$, then we have $a^T\bar{x} = a^Ty$. Now let $y \neq \bar{x}$. The line segment between y and $\bar{x}$ is contained in X. Since $\bar{x} \in \operatorname{rint}(X)$, this line segment can be extended a little bit through the point $\bar{x}$ not leaving X. Hence there exists some point $z \in X$ $\bar{x} = (1-\lambda)y + \lambda z$ for some $\lambda \in [0, 1)$. Since $\bar{x} \neq y$, we have $\lambda \in (0, 1)$. Moreover, $a^Tz \le a^T\bar{x}$. We obtain
$$a^T\bar{x} = (1-\lambda)a^Ty + \lambda a^Tz \le (1-\lambda)a^Ty + \lambda a^T\bar{x}.$$
Hence $(1-\lambda)a^T\bar{x} \le (1-\lambda)a^Ty$. Combined with $a^Ty \le a^T\bar{x}$, we obtain $a^T\bar{x} = a^Ty$. $\square$

We continue preparing our proof of Theorem 1.34. Our next step consists in showing that a singleton and the convex hull of finitely many vectors can be strongly separated, provided that the singleton is not contained in the convex hull.

Lemma 1.36. *Let $x^1, \dots, x^m$ be vectors in $\mathbb{R}^n$ and define $S \coloneqq \operatorname{Conv}(\{x^1, \dots, x^m\})$. If $x \notin S$, then S and $T \coloneqq \{x\}$ can be separated strongly by a hyperplane.*

Proof. The set S can be written as

$$S = \left\{ x \in \mathbb{R}^n \colon \exists \lambda \in \mathbb{R}^m \quad \text{with} \quad \lambda \geq 0, \sum_{i=1}^{m} \lambda_i = 1, x = \sum_{i=1}^{m} \lambda_i x^i \right\}.$$

Hence Theorem 1.30 ensures that S is polyhedrally representable, that is, we can write

$$S = \{ x \in \mathbb{R}^n \colon c_i^T x \leq d_i, \quad i = 1, \ldots, p \},$$

for some $c_i \in \mathbb{R}^n$, $d_i \in \mathbb{R}$, and $p \in \mathbb{N}$. Since $x \notin S$, there exists an index $i \in \{1, \ldots, p\}$ with $c_i^T x > d_i$. Moreover, we have $\sup_{y \in S} c_i^T y \leq d_i$. Combining our estimates, we have

$$\sup_{x \in S} c_i^T y < c_i^T x.$$

Combined with (1.15), we find that the linear form $c_i^T x$ strongly separates S and T. □

Remark 1.37. Let S be a nonempty convex set, such as that considered in Lemma 1.36. If S and $\{0\}$ can be separated, then we have for some nonzero vector $a \in \mathbb{R}^n$,

$$\sup_{x \in S} a^T x \leq 0 \quad \text{and} \quad \inf_{x \in S} a^T x < 0.$$

The set $L \coloneqq \mathrm{Lin}(S)$ is a linear subspace of $\mathbb{R}^n$. Since $a \in \mathbb{R}^n$, we have the unique decomposition $a = a_L + a_{L^\perp}$, where $a_L \in L$, $a_{L^\perp} \in L^\perp$, and $L^\perp = \{ y \in \mathbb{R}^n \colon y^T d = 0 \quad \text{for all} \quad d \in L \}$. Let us show that

$$\sup_{x \in S} a_L^T x \leq 0 \quad \text{and} \quad \inf_{x \in S} a_L^T x < 0.$$

For each $x \in S$, we have $x \in L = \mathrm{Lin}(S)$ and hence $a^T x = a_L^T x + a_{L^\perp}^T x = a_L^T x$. We obtain $\sup_{x \in S} a_L^T x \leq 0$ and $\inf_{x \in S} a_L^T x < 0$. Hence the linear form $a_L^T x$ separates S and $\{0\}$.

Lemma 1.36 provides conditions sufficient to separate a singleton from the convex hull of finitely many points. We would like to extend this separation statement to the separation of a singleton from a convex set which is not necessarily the convex hull of finitely many points. To establish such a separation statement, we cannot directly use Lemma 1.36. However, we can approximate to some extend a nonempty convex set S with the convex hull generated by finitely many points contained in the set S. This approximation is sufficiently accurate if these finitely points in S are close to each point in S.

Lemma 1.38. *If S is a nonempty set in $\mathbb{R}^n$, then there exists a sequence $(x^k) \subset S$ such that each points $x \in S$ is the limit point of a some subsequence of (x^k).*

Lemma 1.38 is a statement from real analysis. It says that any nonempty set in $\mathbb{R}^n$ is *separable*. Using Lemmas 1.36 and 1.38, we can establish the separation of a singleton and a nonempty convex set, provided that the singleton is not contained in the convex set.

Lemma 1.39. *Let S be a nonempty convex set with $0 \notin S$. Then S and $\{0\}$ can be separated.*

Proof. In light of (1.14), we have to establish the existence of a vector $d \in \mathbb{R}^n \setminus \{0\}$ such that

$$\sup_{x \in S} d^T x \leq 0 \quad \text{and} \quad \inf_{x \in S} d^T x < 0.$$

Let $(x^k) \subset S$ be the sequence given by Lemma 1.38. Since S is convex and $(x^k) \subset S$, we have for all $k \in \mathbb{N}$, $\mathrm{Conv}(\{x^1, \ldots, x^k\}) \subset S$. Combined with $0 \notin S$, we have $0 \notin \mathrm{Conv}(\{x^1, \ldots, x^k\})$ for all $k \in \mathbb{N}$. For each $k \in \mathbb{N}$, Lemma 1.36 and remark 1.37 ensure the existence of a vector $a^k \in \mathbb{R}^n \setminus \{0\}$ with $a^k \in \mathrm{Lin}(S)$ with

$$0 > \max_{1 \leq j \leq k} (a^k)^T x^j.$$

Dividing by $\|a^k\|_2$ and defining $d^k := a^k / \|a^k\|_2$, we obtain

$$0 > \max_{1 \leq j \leq k} (d^k)^T x^j.$$

The vectors d^k have norm one and hence there exists a subsequence of (d^{k_ℓ}) of (d^k) such that $d^k \to d$ as $\ell \to \infty$ and we have $\|d\|_2 = 1$. Moreover, $d^k \in \mathrm{Lin}(S)$ for all $k \in \mathbb{N}$. Since $\mathrm{Lin}(S)$ is closed, we have $d \in \mathrm{Lin}(S)$. Taking limits as $\ell \to \infty$, we obtain for all $j \in \mathbb{N}$,

$$0 \geq d^T x^j.$$

Since each $x \in S$ is the limit point of some subsequence of (x^k), we obtain $d^T x \leq 0$ for all $x \in S$. Hence $\sup_{x \in S} d^T x \leq 0$.

In light of (1.14), it must yet be shown that $\inf_{x \in S} d^T x < 0$. Suppose that $\inf_{x \in S} d^T x \geq 0$. Combined with $\sup_{x \in S} d^T x \leq 0$, we obtain $d^T x = 0$ for all $x \in S$ and hence $d^T y = 0$ for all $y \in \mathrm{Lin}(S)$. Combined with $d \in \mathrm{Lin}(S)$, we obtain $0 = d^T d = 1$, a contradiction. □

Now we separate two convex nonempty disjoint sets.

Lemma 1.40. *If S and T are two nonempty convex disjoint sets, then they can be separated by a hyperplane.*

Proof. We apply Lemma 1.39 to the sets $\widehat{S} := S - T$ and $\widehat{T} := \{0\}$. Recall from Proposition 1.7 that the set $\widehat{S} := S - T$ consists of all points $z = x - y$ with $x \in S$ and $y \in T$. Proposition 1.7 ensures that $\widehat{S}$ is a convex set. Since S and T are disjoint, $\widehat{S}$ does not contain the zero vector. Hence Lemma 1.39 ensures that $\widehat{S}$ and $\{0\}$ can be separated by a hyperplane. Hence there exists a nonzero vector $a \in \mathbb{R}^n$ with

$$\sup_{z \in \widehat{S}} a^T z \leq 0 \quad \text{and} \quad \inf_{z \in \widehat{S}} a^T z < 0.$$

Each point $z \in \widehat{S}$ can be written as $z = x - y$ with $x \in S$ and $y \in T$. We obtain

$$\sup_{x \in S, y \in T} a^T x - a^T y \leq 0 \quad \text{and} \quad \inf_{x \in S, y \in T} a^T x - a^T y < 0.$$

Hence

$$\sup_{x \in S} a^T x \leq \inf_{y \in T} a^T y \quad \text{and} \quad \inf_{x \in S} a^T x < \sup_{y \in T} a^T y.$$

This relation is (1.14) and hence S and T can be separated. □

We are now ready to establish Theorem 1.34.

Proof of Theorem 1.34. Let S and T be convex sets in $\mathbb{R}^n$. We show that S and T can be separated by a hyperplane if and only if their relative interiors have no point in common.

Let S and T be separated by a hyperplane. Using (1.14), we have for some nonzero vector $a \in \mathbb{R}^n$,

$$\sup_{x \in S} a^T x \leq \inf_{y \in T} a^T y \quad \text{and} \quad \inf_{x \in S} a^T x < \sup_{y \in T} a^T y.$$

Suppose that the relative interiors of S and T have a point, say $\bar{x}$, in common. Let us show that $\bar{x}$ is a maximizer of $a^T x$ over S. Since $\bar{x} \in T$, we have $\sup_{x \in S} a^T x \leq a^T \bar{x}$. Hence $\bar{x}$ is a maximizer of $a^T x$ over S. Similarly, we can show that $\bar{x}$ is a minimizer of $a^T x$ over T. Lemma 1.35 ensures that $a^T x$ is constant over the union $S \cup T$. This contradicts the separation of S and T.

Now let the relative interiors of S and T be disjoint. We apply Lemma 1.40 to the set $S' \coloneqq \operatorname{rint}(S)$ and $T' \coloneqq \operatorname{rint}(T)$. These two sets are disjoint and Theorem 1.16 says that they are convex and nonempty. Using Lemma 1.40 and (1.14), we have for some nonzero vector $a \neq 0$,

$$\sup_{x \in S'} a^T x \leq \inf_{y \in T'} a^T y \quad \text{and} \quad \inf_{x \in S'} a^T x < \sup_{y \in T'} a^T y. \tag{1.17}$$

We claim that the linear form $a^T x$ also separates S and T. The quantities suprema and infima (1.17) remain unchanged when replacing S' with its closure $\operatorname{cl}(S')$ and T' with its closure $\operatorname{cl}(T')$.[1] Proposition 1.17 yields $\operatorname{cl}(S') = \operatorname{cl}(S)$ and $\operatorname{cl}(T') = \operatorname{cl}(T)$. We also have $T \subset \operatorname{cl}(T)$ and $S \subset \operatorname{cl}(S)$. Hence (1.17) yields

$$\sup_{x \in S} a^T x \leq \inf_{y \in T} a^T y.$$

[1] These assertions can be deduced from the following fact. Let $X \subset \mathbb{R}^n$ be a nonempty set and let $f : \operatorname{cl}(X) \to \mathbb{R}$ be a continuous function. Then

$$\inf_{x \in X} f(x) = \inf_{x \in \operatorname{cl}(X)} f(x).$$

Since $X \subset \operatorname{cl}(X)$, we have $\inf_{x \in X} f(x) \geq \inf_{x \in \operatorname{cl}(X)} f(x)$. For the reverse inequality, we make use of the definition of the infimum.

Suppose that $\inf_{x \in \operatorname{cl}(X)} f(x) > -\infty$. Fix $\varepsilon > 0$. Since $\inf_{x \in \operatorname{cl}(X)} f(x) > -\infty$, there exists $\bar{x} \in \operatorname{cl}(X)$ with $f(\bar{x}) < \inf_{x \in \operatorname{cl}(X)} f(x) + \varepsilon$. Since $x \in \operatorname{cl}(X)$, there exists a sequence $(x^k) \subset X$ with $x^k \to \bar{x}$ as $k \to \infty$. Combined with the fact that f is continuous, we find that $f(x^k) < f(\bar{x}) + \varepsilon$, valid for all

We further have $T' \subset T$ and $S' \subset T$. Therefore, the second inequality in (1.17) yields

$$\inf_{x\in S} a^T x < \sup_{y\in T} a^T y.$$

□

1.9.2 Strong Separation Theorem

We provide conditions necessary and sufficient for strong separation of two convex nonempty sets.

Theorem 1.41. *Two nonempty convex sets S and T in $\mathbb{R}^n$ can be separated strongly by a hyperplane if and only if they have a positive distance to each other, that is,*

$$\inf_{x\in S, y\in T} \|x-y\|_2 > 0.$$

Proof. Suppose that the hyperplane $H(a,b)$ strongly separates S and T with $a \in \mathbb{R}^n\setminus\{0\}$ and $b \in \mathbb{R}$. This means that we can find two numbers $b' < b < b''$ with $S \subset H_-(a,b')$ and $T \subset H_+(a,b'')$. Let us show that S and T must have positive distance. If the distance between the sets is zero, then there exist sequences $(x^k) \subset S$ and $(y^k) \subset T$ with $\|x^k - y^k\|_2 \to 0$ as $k \to \infty$. Hence $a^T(y^k - x^k) \to 0$ as $k \to \infty$. We also have $a^T x^k \le b' < b < b'' \le a^T y^k$ for all $k \in \mathbb{N}$ and hence $a^T(y^k - x^k) \ge a^T y^k - b' \ge b'' - b' > 0$, contradicting the fact that $a^T(x^k - y^k)$ converges to zero.

Now let S and T be nonempty and convex with positive distance to each other. Let us define $\delta \coloneqq \inf_{x\in S, y\in T} \|x-y\|_2$. We consider the set

$$S' \coloneqq S + \{z \in \mathbb{R}^n \colon \|z\|_2 \le \delta/2\}.$$

The set S' is convex according to Proposition 1.7. Moreover T is convex, and S' and T do not intersect. Thus they can be separated by a hyperplane and we have for some nonzero $a \in \mathbb{R}^n$,

$$\sup_{x'\in S'} a^T x' \le \inf_{y\in T} a^T y.$$

Combined with

$$\sup_{x'\in S'} a^T x' = \sup_{x\in S, \|z\|_2\le\delta/2} a^T(x+z) = \sup_{x\in S} a^T + (\delta/2)\|a\|_2,$$

sufficiently large k. Putting together the pieces, we find that

$$\inf_{x\in X} f(x) \le f(x^k) \le f(\bar{x}) + \varepsilon \le \inf_{x\in \mathrm{cl}(X)} f(x) + 2\varepsilon.$$

Since this holds true for all $\varepsilon > 0$, we obtain $\inf_{x\in X} f(x) \le \inf_{x\in\mathrm{cl}(X)} f(x)$.

Suppose that $\inf_{x\in\mathrm{cl}(X)} f(x) = -\infty$. Then for all $k \in \mathbb{N}$ there exists $x^k \in \mathrm{cl}(X)$ with $f(x^k) < -k$. Since $x^k \in \mathrm{cl}(X)$ and f is continuous, we can find a point $y^k \in X$ (close to x^k) with $f(y^k) \le f(x^k)+1$. Hence $f(y^k) \le -k+1$ yielding $\inf_{x\in X} f(x) = -\infty$.

we conclude that

$$\sup_{x\in S} a^T x < \inf_{y\in T} a^T y.$$

Combined with (1.15), we find that S and T can be separated strongly. $\square$

Using Theorem 1.41, we can show that each nonempty closed convex set $X \subset \mathbb{R}^n$ is the intersection of closed halfspaces containing X.

Lemma 1.42. *If $X \subset \mathbb{R}^n$ is nonempty, closed, and convex, then X can be represented as the intersection of closed halfspaces:*

$$X = \bigcap_{H(a,b) \text{ is a hyperplane with } X\subset H_-(a,b)} H_-(a,b).$$

Proof. Since we intersect all closed halfspaces $H_-(a,b)$ with $H(a,b)$ being a hyperplane such that $X \subset H_-(a,b)$, we obtain the inclusion "$\subset$."

Let us now show the reverse inclusion. Let $y \notin X$. Then the sets X and $\{y\}$ have positive distance to each other, as X is closed. Hence Theorem 1.41 ensures the existence of a nonzero vector $a \in \mathbb{R}^n$ and a scalar $b \in \mathbb{R}$ with

$$a^T y > b, \quad \text{and} \qquad a^T x \le b, \quad \text{for all} \quad x \in X.$$

Hence $y \notin H_-(a,b)$. In particular, y is not contained in the intersection over all closed halfspaces containing X. $\square$

1.10 Supporting Hyperplanes

The separation theorem, Theorem 1.41, ensures that a closed and nonempty convex set X is the intersection of all closed halfspaces containing X (see Lemma 1.42). Among these halfspaces, the most interesting are the "extreme" ones — those with the boundary hyperplane touching X.

We recall that the relative boundary of a set $X \subset \mathbb{R}^n$ is the set $\mathrm{cl}(X) \setminus \mathrm{rint}(X)$.

Definition 1.43. *Let $X \subset \mathbb{R}^n$ be a* closed *convex set, and let $\bar{x}$ be a point from its relative boundary. For $a \in \mathbb{R}^n \setminus \{0\}$, the hyperplane*

$$\{x \in \mathbb{R}^n \colon\ a^T x = a^T \bar{x}\}$$

is called supporting *to X at $\bar{x}$ if it separates X and $\{\bar{x}\}$.*

Let X be a closed convex set ant $\bar{x}$ be a point from its relative boundary. Then a hyperplane $H(a,b) = \{x \in \mathbb{R}^n \colon a^T x = b\}$ with $a \neq 0$ supports X at $\bar{x}$ if and only if the linear form $a^T x$ attains its maximum over X, the maximum is equal to b, and the linear form $a^T x$ is nonconstant over X.

Let us discuss the existence of supporting hyperplanes.

Proposition 1.44. *Let $X \subset \mathbb{R}^n$ be a closed convex set and let $\bar{x}$ be a point from its relative boundary. Then*

1. *there exists at least one hyperplane supporting to X at $\bar{x}$, and*
2. *if a hyperplane Π is supporting to X at $\bar{x}$, then the set $X \cap \Pi$ has dimension less than that of X.*

Proof. 1. Since $\bar{x}$ is a point from the relative boundary of X, we have $\bar{x} \notin \operatorname{rint}(X)$. Moreover, we have $\operatorname{rint}(\{\bar{x}\}) = \{\bar{x}\}$. Hence $\operatorname{rint}(\{\bar{x}\}) \cap X = \emptyset$. Therefore, the sets $\{\bar{x}\}$ and X can be separated by a hyperplane. Hence, there exists $a \in \mathbb{R}^n$ with $a \neq 0$ and $b \in \mathbb{R}$ such that

$$a^T x \leq b \quad \text{for all} \quad x \in X \quad \text{and} \quad a^T \bar{x} \geq b.$$

We also have $\bar{x} \in X$. Hence $a^T \bar{x} \leq b \leq a^T \bar{x}$, yielding $b = a^T \bar{x}$.

2. Let $\Pi = \{x \in \mathbb{R}^n \colon\ a^T x = a^T \bar{x}\}$ be a hyperplane supporting X at $\bar{x}$. Then the set $X' := X \cap \Pi$ is nonempty, as it contains $\bar{x}$. Moreover, it is convex as it is the intersection of convex sets. The linear form $a^T x$ is constant on X' and therefore it is also constant on the affine hull $\operatorname{Aff}(X')$. At the same time, the linear form is nonconstant on X by definition of a supporting hyperplane. Thus the affine space $\operatorname{Aff}(X')$ is a proper subset of $\operatorname{Aff}(X)$ and therefore the affine dimension of $\operatorname{Aff}(X')$ is less than that of $\operatorname{Aff}(X)$. (Note that if $P \subset Q$ are two affine subspaces in $\mathbb{R}^n$, then the affine dimension of P is less than or equal to that of Q and the affine dimensions are equal if and only if $P = Q$.)

□

1.11 Exercises

Exercise 1.1.
Which of the following sets are convex? No justifications are required.

1. $\{x \in \mathbb{R}^2 \colon x_1 + i^2 x_2 \leq 1,\ i = 1, \ldots, 10\}$
2. $\{x \in \mathbb{R}^2 \colon x_1^2 + 2ix_1x_2 + i^2x_2^2 \leq 1,\ i = 1, \ldots, 10\}$
3. $\{x \in \mathbb{R}^2 \colon x_1^2 + ix_1x_2 + i^2x_2^2 \leq 1,\ i = 1, \ldots, 10\}$
4. $\{x \in \mathbb{R}^2 \colon x_1^2 + 5x_1x_2 + 4x_2^2 \leq 1\}$
5. $\{x \in \mathbb{R}^{10} \colon x_1^2 + 2x_2^2 + \cdots + 10x_{10}^2 \leq 2004x_1 - 2003x_2 + 2002x_3 - \cdots + 1996x_9 - 1995x_{10}\}$
6. $\{x \in \mathbb{R}^2 \colon \exp(x_1) \leq x_2\}$
7. $\{x \in \mathbb{R}^2 \colon \exp(x_1) \geq x_2\}$
8. $\{x \in \mathbb{R}^n \colon \sum_{i=1}^n x_i^2 = 1\}$
9. $\{x \in \mathbb{R}^n \colon \sum_{i=1}^n x_i^2 \leq 1\}$
10. $\{x \in \mathbb{R}^n \colon \sum_{i=1}^n x_i^2 \geq 1\}$
11. $\{x \in \mathbb{R}^n \colon \max_{1 \leq i \leq n} x_i \leq 1\}$
12. $\{x \in \mathbb{R}^n \colon \max_{1 \leq i \leq n} x_i \geq 1\}$
13. $\{x \in \mathbb{R}^n \colon \max_{1 \leq i \leq n} x_i = 1\}$
14. $\{x \in \mathbb{R}^n \colon \min_{1 \leq i \leq n} x_i \leq 1\}$
15. $\{x \in \mathbb{R}^n \colon \min_{1 \leq i \leq n} x_i \geq 1\}$
16. $\{x \in \mathbb{R}^n \colon \min_{1 \leq i \leq n} x_i = 1\}$

Exercise 1.2 (see [7, Exercise 2.2])**.**
Show that a set is convex if and only if its intersection with any line is convex.

Exercise 1.3 (Expanded and restricted sets, see [7, Exercise 2.14])**.**
Let $S \subset \mathbb{R}^n$, and let $\|\cdot\|$ be a norm on $\mathbb{R}^n$.

1. For $a \geq 0$, we define $S_a = \{x \in \mathbb{R}^n \colon \operatorname{dist}(x, S) \leq a\}$, where $\operatorname{dist}(x, S) = \inf_{y \in S} \|x - y\|$. We refer to S_a as S *expanded* or *extended* by a. Show that if S is nonempty and convex, then S_a is convex.

2. For $a \geq 0$, we define $S_{-a} = \{x \in \mathbb{R}^n : B(x,a) \subset S\}$, where $B(x,a)$ is the closed ball (in the norm $\|\cdot\|$) centered at x with radius a, that is, $B(x,a) = \{y \in \mathbb{R}^n : \|y-x\| \leq a\}$. We refer to S_{-a} as S *shrunk* or *restricted* by a, since S_{-a} consists of all points that are at least a distance a from $\mathbb{R}^n \setminus S$. Show that if S is convex, then S_{-a} is convex.

Exercise 1.4 (A set of hyperplanes, see [7, Exercise 2.21]).
Suppose that C and D are disjoint sets in $\mathbb{R}^n$. Consider the set of points $(a,b) \in \mathbb{R}^{n+1}$ for which $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$. Show that this set is a cone.
Hint: Use Example 1.5.

Exercise 1.5 (see [7, Exercise 2.12]).
Which of the following sets are convex?

1. A slap, that is, a set of the form $\{x \in \mathbb{R}^n : \alpha \leq a^T x \leq \beta\}$.
2. A rectangle, that is, a set of the form $\{x \in \mathbb{R}^n : \alpha_i \leq x_i \leq \beta_i,\ i = 1, \ldots, n\}$.
3. A wedge, that is, $\{x \in \mathbb{R}^n : a_1^T x \leq b_1,\ a_2^T x \leq b_2\}$.
4. The set of points closer to a given point than a given set, that is,
$$\{x \in \mathbb{R}^n : \|x - x_0\|_2 \leq \|x - y\|_2 \quad \text{for all} \quad y \in S\},$$
where $S \subset \mathbb{R}^n$.
5. The set of points closer to one set than another, that is,
$$\{x \in \mathbb{R}^n : \operatorname{dist}(x,S) \leq \operatorname{dist}(x,T)\},$$
where S, $T \subset \mathbb{R}^n$, and
$$\operatorname{dist}(x,S) = \inf_{z \in S} \|x - z\|_2.$$
6. The set $\{x \in \mathbb{R}^n : x + S_2 \subset S_1\}$, where S_1, $S_2 \subset \mathbb{R}^n$ and S_1 is convex.
7. The set of points whose distance to a does not exceed fixed fraction θ of the distance to b, that is, the set $\{x \in \mathbb{R}^n : \|x - a\|_2 \leq \theta \|x - b\|_2\}$. You can assume $a \neq b$ and $\theta \in [0,1]$.

Exercise 1.6.
Which of the following statements are true? For true statements, provide proofs; for wrong statements, provide counterexamples.

1. The convex hull of a closed set in $\mathbb{R}^n$ is closed.

2. The convex hull of a closed convex set in $\mathbb{R}^n$ is closed.

3. The convex hull of a closed, bounded set in $\mathbb{R}^n$ is closed and bounded.

Hints: (i) Use the fact that a set in $\mathbb{R}^n$ is compact if and only if it is bounded and closed. (ii) Continuous functions map compact sets to compact ones.

Exercise 1.7 (see [7, Exercise 2.16]).
Show that if S_1 and S_2 are convex sets in $\mathbb{R}^{m\times n}$, then their partial sum

$$S = \{(x, y_1 + y_2) \colon x \in \mathbb{R}^m,\ y_1, y_2 \in \mathbb{R}^n,\ (x, y_1) \in S_1,\ (x, y_2) \in S_2\}$$

is convex.

Exercise 1.8 (see [3, Exercise 1.2]).
Establish the following statements.

1. The intersection $\cap_{i\in I} C_i$ of a collection C_i, $i \in I$, of cones is a cone.
2. The Cartesian product $C_1 \times C_2$ of two cones C_1 and C_2 is a cone.
3. The sum $C_1 + C_2$ of two conic sets C_1 and C_2 is conic.
4. If $C \subset \mathbb{R}$ is a cone and $A \in \mathbb{R}^{m\times n}$ is a matrix, then $A(C) \coloneqq \{Ax \colon x \in C\}$ and $A^{-1}(C) \coloneqq \{x \in \mathbb{R}^n \colon Ax \in C\}$ are cones.

Exercise 1.9 (Taking the image of an affine mapping).
Let $X \subset \mathbb{R}^n$ be convex, let $A \in \mathbb{R}^{m\times n}$, and let $b \in \mathbb{R}^m$. Let us define the mapping $\mathcal{A}$ by $\mathcal{A}(x) \coloneqq Ax + b$. Show that the image of X under the mapping $\mathcal{A}$,

$$\mathcal{A}(X) \coloneqq \{Ax + b \colon x \in X\},$$

is convex.

Exercise 1.10 (Taking the inverse image under affine mapping).
Let $X \subset \mathbb{R}^n$ be convex, let $A \in \mathbb{R}^{n\times m}$, and let $b \in \mathbb{R}^n$. Let us define the mapping $\mathcal{A}$ by $\mathcal{A}(y) = Ay + b$. Show that the inverse image of X under the mapping $\mathcal{A}$,

$$\mathcal{A}^{-1}(X) \coloneqq \{y \in \mathbb{R}^m \colon Ay + b \in X\},$$

is convex.

Exercise 1.11.
Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite and let $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Show that if $b^T A^{-1} b - c > 0$, then the set

$$X = \{x \in \mathbb{R}^n \colon x^T Ax + 2b^T x + c \leq 0\}$$

is an ellipsoid. (If $c \leq 0$ and $b \neq 0$ or if $c < 0$ and $b = 0$, then $b^T A^{-1} b - c > 0$).

Exercise 1.12 (Kirchberger's Theorem)**.**
Prove the following Kirchberger's Theorem:

Let $X = \{x^1, \dots, x^k\}$ and $Y = \{y^1, \dots, y^m\}$ be finite subsets in $\mathbb{R}^n$ with and $k+m \geq n+2$. Let the points $x^1, \dots, x^k$, $y^1, \dots, y^m$ be distinct. If for each subset $S \subset X \cup Y$ with $n+2$ points the convex hulls of the sets $X \cap S$ and $Y \cap S$ are disjoint, then convex hulls of X and Y are disjoint.

Hint: Either one of the following approaches might be helpful:

1. Assume on the contrary that the convex hulls of X and Y intersect so that
$$\sum_{i=1}^{k} \lambda_i x^i = \sum_{j=1}^{m} \mu_j y^j$$
for certain nonnegative numbers $\lambda_i \geq 0$, $\mu_j \geq 0$ with $\sum_{i=1}^{k} \lambda_i = 1$ and $\sum_{j=1}^{p} \mu_j = 1$, and look at this expression with the minimum total number of nonzero coefficients λ_i, μ_j.

2. Show that if the convex hulls of X and Y intersect, then there exists a set $T \subset X \cup Y$ with at most $n+2$ elements such that the convex hulls of the sets $X \cap T$ and $Y \cap T$ intersect. Subsequently, deduce Kirchberger's theorem.

Exercise 1.13.
Exercise 1.12 motivates the following statement:

Let X and Y be subsets of $\mathbb{R}^n$. If the convex hulls of X and Y intersect, then there exists a set $S \subset X \cup Y$ with at most $n+2$ elements such that the convex hulls of the sets $X \cap S$ and $Y \cap S$ intersect.

Establish this statement.

Exercise 1.14 (Shapley–Folkman theorem)**.**
The following statement is known as the Shapley–Folkman theorem.

Let $X_i \subset \mathbb{R}^n$, $i = 1, \dots, m$, be nonempty sets with $m \geq n$. Define $X := X_1 + \dots + X_m$. Then each $x \in \operatorname{Conv}(X)$ has a representation $x = \sum_{i=1}^{m} x^i$ such that $x^i \in \operatorname{Conv}(X_i)$ for all indices $i \in \{1, \dots, m\}$, and $x^i \in X_i$ for at least $(m-n)$ indices i.

Establish the Shapley–Folkman theorem.

Exercise 1.15 (A min-max optimization problem)**.**
Let $f_i\colon \mathbb{R}^n \to \mathbb{R}$ be convex functions for $i = 1, \dots, N$ with $N \geq n + 1$ and $N \in \mathbb{N}$. We consider

$$\inf_{x \in \mathbb{R}^n} \max_{i \in \{1,\dots,N\}} f_i(x).$$

Let $f^* \in \mathbb{R}$ be its optimal value. Show that there exists an index set $I \subset \{1, \dots, N\}$ with at most $n + 1$ elements such that the optimal value of

$$\inf_{x \in \mathbb{R}^n} \max_{i \in I} f_i(x).$$

equals f^*.

Hint: Consider the sets $X_i := \{x \in \mathbb{R}^n \colon f_i(x) < f^*\}$, $i = 1, \dots, N$, and use Helly's theorem.

Exercise 1.16.
Let A, $B \subset \mathbb{R}^n$ be nonempty sets. Show that $\operatorname{Conv}(A + B) = \operatorname{Conv}(A) + \operatorname{Conv}(B)$.

Exercise 1.17 (**Multivariate Chernoff-type bound [5, Thms. 1.3.1 and 1.3.2], [6, Thm. 1.2])**.**
Let ξ be a random vector with values in $\mathbb{R}^n$, and let $B \subset \mathbb{R}^n$ be a (measurable) convex set. Show that

$$\operatorname{Prob}(\xi \in B) \leq \mathrm{e}^{-\inf_{\zeta \in B} \sup_{a \in \mathbb{R}^n} \{\, a^T\zeta - \ln \mathbb{E}[\mathrm{e}^{a^T\xi}] \,\}}.$$

Exercise 1.18 (Separation of sets)**.**
Which of the following statements are true? For true statements, provide proofs; for wrong statements, provide simple counterexamples.

1. If X and Y are nonempty disjoint sets, then they can be strongly separated by a hyperplane.
2. If X and Y are nonempty disjoint sets and their interiors do not intersect, then X and Y can be separated by a hyperplane.
3. Let $X \subset \mathbb{R}^n$ be nonempty, *closed*, and convex. If $y \notin X$, then there exist $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$ such that $a^T y > b \geq a^T x$ for all $x \in X$.

Exercise 1.19 (Homogeneous Farkas lemma).
Establish Farkas' lemma:

If $a_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, and $a \in \mathbb{R}^n$, then exactly of the following two systems has a solution: (i) $\lambda \in \mathbb{R}^m$ *with* $a = \sum_{i=1}^m \lambda_i a_i$ *and* $\lambda_i \geq 0$, $i = 1, \ldots, m$. (ii) $x \in \mathbb{R}^n$ *such that* $a_i^T x \leq 0$, $i = 1, \ldots, m$, *and* $a^T x > 0$.

Hint: For one implication, apply Theorem 1.41 to $\operatorname{Cone}(\{a_1, \ldots, a_m\})$ and $\{a\}$. Is $\operatorname{Cone}(\{a_1, \ldots, a_m\})$ closed?

Exercise 1.20 (Separation Properties of Cones [3, Problem 1.50]).
Define a homogeneous halfspace to be a closed halfspace associated with a hyperplane that passes through the origin. Show that:

1. A nonempty closed convex cone is the intersection of the homogeneous halfspaces that contain it.
2. The closure of the convex cone generated by a nonempty set X is the intersection of all the homogeneous halfspaces containing X.

Exercise 1.21 (Characterization of Closed Convex Sets [3, Problem 1.52]*).
Let C be a nonempty closed convex subset of $\mathbb{R}^{n+1}$. Show that if C contains no vertical lines, then C is the intersection of the closed halfspaces that contain it and correspond to nonvertical hyperplanes.

Hints:

1. A hyperplane $\{x \in \mathbb{R}^{n+1} \colon a^T x = b\}$ with $a \in \mathbb{R}^{n+1} \setminus \{0\}$ and $b \in \mathbb{R}$ is called *nonvertical* if $a_{n+1} \neq 0$.
2. Let $w \in \mathbb{R}^n$. The set $\{(w, \tau) \colon \tau \in \mathbb{R}\}$ is called *vertical line* in $\mathbb{R}^{n+1}$.
3. Use Proposition 1.5.8 in [3].

Exercise 1.22 (A polyhedral set).
Let p, $q \in \mathbb{N}$. Let $\{x^1, \ldots, x^p\} \subset \mathbb{R}^n$, and let $\{y^1, \ldots, y^q\} \subset \mathbb{R}^n$. We define

$$X \coloneqq \operatorname{Conv}(\{x^1, \ldots, x^p\}) + \operatorname{Cone}(\{y^1, \ldots, y^q\}).$$

For $m \in \mathbb{N}$, we define $\mathbb{R}^m_+ \coloneqq \{x \in \mathbb{R}^m \colon x_i \geq 0, \quad i = 1, \ldots, m\}$.

1. Show that

$$X = \left\{ x \in \mathbb{R}^n \colon \quad \exists (\lambda, \mu) \in \mathbb{R}^p_+ \times \mathbb{R}^q_+, \quad \sum_{i=1}^p \lambda_i = 1, \quad x = \sum_{i=1}^p \lambda_i x^i + \sum_{j=1}^q \mu_j y^j \right\}.$$

2. Show that X is polyhedral. Provide a concise answer.

Exercise 1.23 (Minkowski sums of polyhedral sets)**.**
Given two polyhedral sets, is their Minkowski sum a polyhedral set?

2 Convex Functions

Definition 2.1. *A function $f : X \to \mathbb{R}$ defined on a nonempty set $X \subset \mathbb{R}^n$ is called* convex *if*

1. *X is convex and*
2. *for all x, $y \in X$ and all $\lambda \in [0, 1]$,*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

A function $f : X \to \mathbb{R}$ is called *concave* if $-f : X \to \mathbb{R}$ is a convex function. Figure 2.1 provides an illustration.

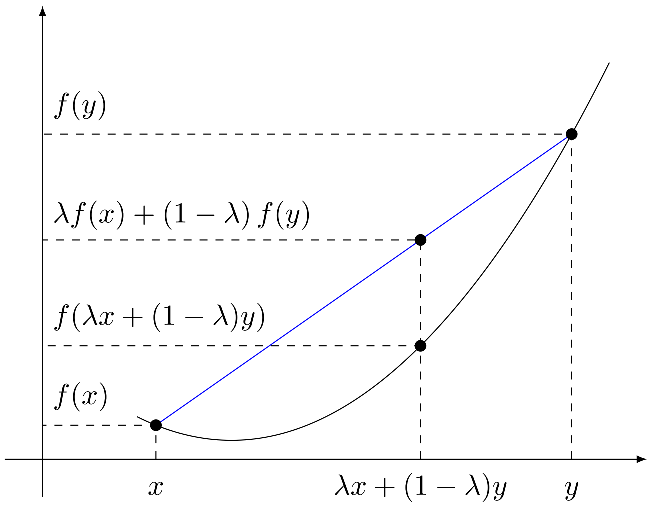

Figure 2.1: Illustration of the definition of a convex function. The graph of $[0, 1] \ni \lambda \mapsto \lambda f(x) + (1 - \lambda)f(y)$ is above that of $[0, 1] \ni \lambda \mapsto f(\lambda x + (1 - \lambda)y)$.

Example 2.2.

1. Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The affine function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = a^T x + b$ is both convex and concave.
2. Each norm $\| \cdot \|$ on $\mathbb{R}^n$ is a convex function. Let $\lambda \in [0, 1]$ and x, $y \in \mathbb{R}^n$. Using the triangle inequality, we obtain $\|\lambda x + (1 - \lambda)y\| \leq \|\lambda x\| + \|(1 - \lambda)y\| = \lambda\|x\| + (1 - \lambda)\|y\|$.
3. For $X = \mathbb{R}$, the functions $x \mapsto \exp(x)$ and $x \mapsto x^{2p}$ for $p \in \mathbb{N}$ are convex. The convexity of these function can be established using Theorem 2.9.
4. For $X = \mathbb{R}_+$, the functions $x \mapsto x^p$ for $p \in [1, \infty)$ and $x \mapsto -x^p$ for $p \in [0, 1]$ are convex. The convexity of these function can be established using Theorem 2.9.

Let $X \subset \mathbb{R}^n$ be nonempty and let $f : X \to \mathbb{R}$. The set

$$\operatorname{epi}(f) := \{(x,t) \in X \times \mathbb{R} \colon\ f(x) \le t\}.$$

is called *epigraph* of f.

Theorem 2.3. *If $X \subset \mathbb{R}^n$ is nonempty and convex, then $f : X \to \mathbb{R}$ is convex if and only if* $\operatorname{epi}(f)$ *is convex.*

Proof. See Exercise 2.1. □

2.1 Jensen's inequality

Below we state the important Jensen inequality.

Proposition 2.4. *Let $X \subset \mathbb{R}^n$ be nonempty and convex, and let $f : X \to \mathbb{R}$ be convex. Then for all $x_i \in X$, $\lambda_i \ge 0$, $i = 1, \ldots, m$, with $\sum_{i=1}^m \lambda_i = 1$,*

$$f\Bigg(\sum_{i=1}^m \lambda_i x_i\Bigg) \le \sum_{i=1}^m \lambda_i f(x_i).$$

Proof. The points $(x_i, f(x_i))$ are contained in $\operatorname{epi}(f)$. Since $\operatorname{epi}(f)$ is convex according to Theorem 2.3, and $\sum_{i=1}^m \lambda_i x_i$ and $\sum_{i=1}^m \lambda_i f(x_i)$ are convex combinations, we have

$$\Bigg(\sum_{i=1}^m \lambda_i x_i, \sum_{i=1}^m \lambda_i f(x_i)\Bigg) \in \operatorname{epi}(f).$$

Using the definition of $\operatorname{epi}(f)$, we obtain Jensen's inequality. □

Jensen's inequality can be generalized. For example, if $f : \mathbb{R}^n \to \mathbb{R}$ is convex and $\xi \in \mathbb{R}^n$ is an integrable random vector, then $f(\mathbb{E}[\xi]) \le \mathbb{E}[f(\xi)]$.

Example 2.5. Let $p_1, \ldots, p_n > 0$ with $\sum_{i=1}^n p_i = 1$ and $q_1, \ldots, q_n > 0$ with $\sum_{i=1}^n q_i = 1$. We show that the Kullback–Leibler distance

$$\sum_{i=1}^n p_i \ln\left(\frac{p_i}{q_i}\right)$$

between p and q is nonnegative.

The function $f(x) = -\ln(x)$ on $\{x \in \mathbb{R} \colon x > 0\}$ is convex. Defining $x_i = q_i/p_i$ and $\lambda_i = p_i$, we have

$$0 = -\ln\left(\sum_{i=1}^n q_i\right) = f\left(\sum_{i=1}^n p_i x_i\right) \le \sum_{i=1}^p p_i f(x_i) = -\sum_{i=1}^n p_i \ln(q_i/p_i) = \sum_{i=1}^n p_i \ln(p_i/q_i).$$

Example 2.6. Jensen' inequality can be used to establish the inequality of arithmetic and geometric means

$$(x_1 \cdot x_2 \cdots x_n)^{1/n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n} \quad \text{for all} \quad x \in \mathbb{R}^n_+.$$

Let us establish this inequality using Jensen's inequality. If one component of x is equal to zero, then this inequality holds true. Let now $x > 0$. The function $f(t) = -\ln(t)$ is convex on $(0, \infty)$. Jensen's inequality implies

$$\begin{aligned} -\ln\left(\frac{\sum_{i=1}^n x_i}{n}\right) = f\left(\frac{\sum_{i=1}^n x_i}{n}\right) &\leq (1/n)\sum_{i=1}^n f(x_i) = -(1/n)\sum_{i=1}^n \ln(x_i) \\ &= -\sum_{i=1}^n \ln(x_i^{1/n}) = -\ln(x_1^{1/n} \cdots x_n^{1/n}). \end{aligned}$$

Multiplying the inequality with -1 and applying the exponential function, we obtain the inequality of arithmetic and geometric means.

2.2 Extended real-valued functions

By convention, it is convenient to think that a convex function f is defined *everywhere* on $\mathbb{R}^n$ and takes real values *and* the value $+\infty$. With this convention, we say that a convex function f on $\mathbb{R}^n$ is a function taking values in the extended reals $\mathbb{R} \cup \{+\infty\}$ such that its domain $\operatorname{dom}(f) := \{x \in \mathbb{R}^n : f(x) \in \mathbb{R}\}$ is *nonempty* and for all x, $y \in \mathbb{R}^n$ and every $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \tag{2.1}$$

where we use the following rules

$$\begin{aligned} &+\infty \leq +\infty, \\ &\text{if } a \in \mathbb{R}, \text{ then } a + (+\infty) = (+\infty) + (+\infty) = +\infty, \\ &0 \cdot (+\infty) = 0, \\ &\text{if } a > 0, \text{ then } a \cdot (+\infty) = +\infty. \end{aligned}$$

Operations such as $(+\infty) - (+\infty)$ and $(-5) \cdot (+\infty)$ are undefined. The domain $\operatorname{dom}(f)$ of f is the set of points where f is finite.

It is not clear in advance that our new definition of a convex function is equivalent to the initial one: initially we included into the definition requirement for the domain be convex, and now we omit explicit indicating this requirement. In fact, of course, the definitions are equivalent: convexity of $\operatorname{dom}(f)$ is a consequence of (2.1). Indeed, if x, $y \in \operatorname{dom}(f)$ and $\lambda \in [0, 1]$, then the "convexity inequality" (2.1) ensures $\lambda x + (1 - \lambda)y \in \operatorname{dom}(f)$.

It is convenient to think of a convex function as a function defined on $\mathbb{R}^n$, as it saves many technical wordings. For example, with this convention, we can write $f + g$ for two convex functions f and g on $\mathbb{R}^n$ without explicitly indicating the domain of the new function $f+g$. Without this convention, we would need to say $f+g$ is a function with its domain being the intersection of those of f and g, and on this intersection $f+g$ is defined by $(f+g)(x) = f(x) + g(x)$.

An extended real-valued function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called proper if there exists $x \in \mathbb{R}^n$ such that $f(x) < +\infty$. In other words, $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper if its domain is nonempty. This is to say, $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper if it is finite at at least one point.

2.3 Convexity-Preserving Operations

We discuss important operations that preserve convexity functions.

1. **Taking conic combinations:** If $f_i(x)$ are convex functions on $\mathbb{R}^n$ and $\lambda_i \geq 0$ for $i = 1, \ldots, m$, then the function $\sum_{i=1}^m \lambda_i f_i(x)$ is convex, provided that it is finite at at least one point in $\mathbb{R}^n$.

2. **Affine substitutions of arguments:** If $f(x)$ is a convex function on $\mathbb{R}^n$, $A \in \mathbb{R}^{n\times m}$ is a matrix and $b \in \mathbb{R}^n$ is a vector, then the function $g(y) = f(Ay+b)$ is convex on $\mathbb{R}^m$, provided that it is finite at least at one point.

3. **Taking suprema:** If $f_\alpha(x)$ with $\alpha \in \mathcal{A}$ is a family of convex functions on $\mathbb{R}^n$, then the function $\sup_{\alpha\in\mathcal{A}} f_\alpha(x)$ is convex on $\mathbb{R}^n$, provided that it is finite at least at one point.

   This fact can be established by using $\operatorname{epi}(\sup_{\alpha\in\mathcal{A}} f(\cdot)) = \bigcap_{\alpha\in\mathcal{A}} \operatorname{epi}(f_\alpha(\cdot))$ combined with the fact that intersections of convex sets are convex.

   For example, if $\mathcal{A} = \{1, \ldots, m\}$ and f_i are convex functions for $i = 1, \ldots, m$, then $g(x) = \max_{1\leq i\leq m} f_i(x)$ is a convex function, provided that g is finite at least at one point in $\mathbb{R}^n$.

4. **Superposition theorem:** Let $f_i(x)$ be convex functions on $\mathbb{R}^n$ with $i = 1, \ldots, m$ and $F(y_1, \ldots, y_m)$ be convex and *monotonically increasing* on $\mathbb{R}^m$ in the sense that $v \leq w$ implies $F(v) \leq F(w)$. Then the function

   $$g(x) = \begin{cases} F(f_1(x), \ldots, f_m(x)), & \text{if} \quad x \in \bigcap_{i=1,\ldots,m} \operatorname{dom}(f_i), \\ +\infty & \text{otherwise} \end{cases}$$

   is convex, provided that $\bigcap_{i=1,\ldots,m} \operatorname{dom}(f_i)$ is nonempty.

5. **Partial minimization**: Let $f(x, y)$ be a convex function on $(x, y) \in \mathbb{R}^{n\times m}$ and define

   $$g(x) = \inf_{y\in\mathbb{R}^m} f(x, y).$$

   Suppose that $g(x) > -\infty$ for all $x \in \mathbb{R}^n$ and $g(y) < \infty$ for some point $y \in \mathbb{R}^n$. Then g is convex.

Let us verify the convexity of g. By assumption g takes only real values and the value $+\infty$. Let x' and x'' be vectors in $\mathbb{R}^n$ and let us show that for all $\lambda \in [0,1]$,

$$g(\lambda x' + (1-\lambda)x'') \leq \lambda g(x') + (1-\lambda)g(x'').$$

If λ is either 0 or 1, then we have nothing to show. Now let $\lambda \in (0,1)$. If either $g(x') = +\infty$ or $g(x'') = +\infty$, then we have nothing to verify. Let now $g(x')$ and $g(x'')$ be finite. Let $\varepsilon > 0$. Since $g(x') \in \mathbb{R}$, there exists $y' \in \mathbb{R}^m$ such that $f(x', y') \leq g(x') + \varepsilon$. Similarly, there exists y'' with $f(x'', y'') \leq g(x'') + \varepsilon$. We compute

$$\begin{aligned} g(\lambda x' + (1-\lambda)x'') &\leq f(\lambda x' + (1-\lambda)x'', \lambda y' + (1-\lambda)y'') \\ &\leq \lambda f(x', y') + (1-\lambda) f(x'', y'') \\ &\leq \lambda\bigl(g(x') + \varepsilon\bigr) + (1-\lambda)\bigl(g(x'') + \varepsilon\bigr) \\ &= \lambda g(x') + (1-\lambda)g(x'') + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we obtain $g(\lambda x' + (1-\lambda)x'') \leq \lambda g(x') + (1-\lambda)g(x'')$.

6. Projective transformation. Let $f(x)$ be a convex function on $\mathbb{R}^n$. Then the function $F(\alpha, x) = \alpha f(x/\alpha)$ is a convex function on the set $\{\alpha \colon \alpha > 0\} \times \mathbb{R}^n$.

Let us verify the convexity of F. Since f is finite at at least one point $x \in \mathbb{R}$ and $F(1, x) = f(x)$, the function F is finite at the point $(1, x)$. Hence the domain $\mathrm{dom}(F)$ of F is nonempty.

Next, we establish the "convexity inequality" (2.1). Let x, $y \in \mathbb{R}^n$, α, $\beta > 0$, and let $\lambda \in (0,1)$. We need to verify the inequality

$$F(\lambda\alpha + (1-\lambda)\beta, \lambda x + (1-\lambda)y) \leq \lambda F(\alpha, x) + (1-\lambda)F(\beta, y). \tag{2.2}$$

For the right-hand side in (2.2), we obtain

$$\lambda F(\alpha, x) + (1-\lambda)F(\beta, y) = \lambda\alpha f(x/\alpha) + (1-\lambda)\beta f(y/\beta).$$

For the left-hand side in (2.2), we obtain

$$\begin{aligned} &F(\lambda\alpha + (1-\lambda)\beta, \lambda x + (1-\lambda)y) \\ &\quad = (\lambda\alpha + (1-\lambda)\beta) f([\lambda x + (1-\lambda)y]/(\lambda\alpha + (1-\lambda)\beta)) \\ &\quad = (\lambda\alpha + (1-\lambda)\beta) f\left(\frac{\lambda\alpha}{\lambda\alpha + (1-\lambda)\beta}\frac{x}{\alpha} + \frac{(1-\lambda)\beta}{\lambda\alpha + (1-\lambda)\beta}\frac{x}{\alpha}\right) \end{aligned}$$

Combining these expression with the convexity of f, we obtain (2.2).

For example, the function $\alpha \ln(\alpha/\beta)$ is convex on the set $\{(\alpha, \beta) \colon \alpha > 0, \beta > 0\}$. This can be verified by applying the projective transformation to $\ln(1/\beta) = -\ln(\beta)$.

2.4 Derivative-Based Criteria of Convexity

Convexity of sets and functions are one-dimensional properties. A nonempty set $X \subset \mathbb{R}^n$ is convex if and only if the one-dimensional sets

$$\{t \in \mathbb{R} \colon x + th \in X\}$$

are convex for each x, $h \in X$.

Lemma 2.7. *A proper function $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if for each x, $h \in \mathbb{R}^n$, the function*

$$\phi(t) \coloneqq f(x + th)$$

on $\mathbb{R}$ is either convex or identically equal to $+\infty$.

Proof. "⇒" Let us note that each of the functions ϕ is obtained via an affine substitution of f. This is a convexity-preserving operation as discussed in Section 2.3, provided that ϕ is proper. If ϕ is not proper, then it is identically equal to $+\infty$.

"⇐" The set $\operatorname{dom}(f)$ is nonempty because f is proper. Let x, $y \in \operatorname{dom}(f)$, and let $\lambda \in [0, 1]$. We define $h \coloneqq y - x$ and $\phi(t) \coloneqq f(x + t(y - x))$. Since $\phi(0) < +\infty$, ϕ is convex by assumption. Since ϕ is convex, we obtain

$$\phi(\lambda 0 + (1 - \lambda)1) \le \lambda\phi(0) + (1 - \lambda)\phi(1).$$

Combined with $\phi(0) = f(x)$, $\phi(1) = f(y)$, and $\phi(1 - \lambda) = f(x + (1 - \lambda)(y - x)) = f(\lambda x + (1 - \lambda)y)$, we obtain the convexity of f on $\operatorname{dom}(f)$. Hence f is convex on $\mathbb{R}^n$. □

Let us now consider a convex function ϕ on an interval (a, b) and let x, y, $z \in \mathbb{R}$ with $a < x < z < y < b$. Let us assume that the derivatives $\phi'(x)$ and $\phi'(y)$ exist. Our intention is to show that $\phi'(x) \le \phi'(y)$. In other words, ϕ' is monotonically nondecreasing on (a, b), provided that ϕ is convex and differentiable. The point z can be written as a convex combination of x and y:

$$z = \frac{y - z}{y - x} x + \frac{z - x}{y - x} y.$$

Since ϕ is convex, we obtain

$$\phi(z) \le \frac{y - z}{y - x}\phi(x) + \frac{z - x}{y - x}\phi(y).$$

Hence

$$\frac{\phi(z) - \phi(x)}{z - x} \le \frac{\phi(y) - \phi(z)}{y - z} \tag{2.3}$$

Taking limits in (2.3), as $z \to x + 0$, we obtain

$$\phi'(x) \le \frac{\phi(y) - \phi(x)}{y - x}.$$

Similary, taking limits in (2.3) as $z \to y - 0$, we obtain

$$\frac{\phi(y) - \phi(x)}{y - x} \le \phi'(y).$$

Hence, we have $\phi'(x) \le \phi'(y)$. To summarize, if ϕ is convex on an interval (a, b) and differentiable on (a, b), then ϕ' is monotonically nondecreasing.

It turns out that this monotonicity is also sufficient for convexity.

Proposition 2.8. *Let (a, b) be an interval with $-\infty \le a < b \le +\infty$.*

1. *Let ϕ be a differentiable function on (a, b). Then ϕ is convex on (a, b) if and only if its derivative ϕ' is monotonically nondecreasing on (a, b).*
2. *Let ϕ be a twice differentiable function on (a, b). Then ϕ is convex on (a, b) if and only if its second derivative ϕ'' is nonnegative on (a, b).*

Proof. 1. We have already shown that ϕ' is monotonically nondecreasing on (a, b) if ϕ is convex on (a, b).

Let now ϕ' be monotonically nondecreasing on (a, b) and let us show that ϕ is convex on (a, b). Let x, $y \in (a, b)$ with $x < y$ and let $\lambda \in (0, 1)$. We define $z = \lambda x + (1 - \lambda)y$. We must show that

$$\phi(z) \le \lambda\phi(x) + (1 - \lambda)\phi(y).$$

This is the same as

$$\frac{\phi(z) - \phi(x)}{\lambda} \le \frac{\phi(y) - \phi(z)}{1 - \lambda}.$$

Since $z - x = \lambda(y - x)$ and $y - z = (1 - \lambda)(y - x)$, we see that the inequality to be established is equivalent to

$$\frac{\phi(z) - \phi(x)}{z - x} \le \frac{\phi(y) - \phi(z)}{y - z}. \tag{2.4}$$

By the Lagrange mean value theorem, the left-hand side is equal to $\phi'(\xi)$ for some $\xi \in (x, z)$ and the right-hand side is equal to $\phi'(\eta)$ for some $\eta \in (z, y)$. Since ϕ' is nondecreasing and $\xi \le z \le \eta$, we have $\phi'(\xi) \le \phi'(\eta)$. Therefore, we obtain (2.4).

2. We recall from calculus that a differentiable function on (a, b) is monotonically nondecreasing if and only if its derivative is nonnegative on (a, b). We apply this fact to the function ϕ'.

□

Using Proposition 2.8, we can establish an important necessary and sufficient condition for convexity of a smooth function of n variables.

Theorem 2.9. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function. Suppose that the domain* $\mathrm{dom}(f)$ *of f is a convex set with nonempty interior and that f is continuous on* $\mathrm{dom}(f)$ *and twice differentiable on the interior of* $\mathrm{dom}(f)$*. Then f is convex if and only if its Hessian is positive semidefinite on the interior of* $\mathrm{dom}(f)$*:*

$$h^T \nabla^2 f(x) h \ge 0 \quad \textit{for all} \quad h \in \mathbb{R}^n \quad \textit{and for all} \quad x \in \mathrm{int}(\mathrm{dom}(f)).$$

Proof. If f is convex and $x \in \text{int}(\text{dom}(f))$, then the function $\phi(t) = f(x + th)$ for arbitrary $h \in \mathbb{R}^n$ is convex on a neighborhood of the point $t = 0$, as it is the affine substitution of the function f. Since f is twice differentiable in a neighbourhood of x, ϕ is twice differentiable in a neighbourhood of $t = 0$ and we have $\phi''(0) = h^T \nabla^2 f(x) h$. Combined with Proposition 2.8, we find that $h^T \nabla^2 f(x) h \geq 0$.

Now let $h^T \nabla^2 f(x) h \geq 0$ for all $x \in \text{int}(\text{dom}(f))$ and all $h \in \mathbb{R}^n$. Let us first show that f is convex on $\text{int}(\text{dom}(f))$. Proposition 1.14 ensures that $\text{int}(\text{dom}(f))$ is a convex set. Let x, $y \in \text{int}(\text{dom}(f))$. Since convexity is a one-dimensional property, it suffices to show that $\phi(t) = f(x + t(y - x))$ is a convex function on $[0, 1]$. The function ϕ is twice differentiable on $(0, 1)$. Combined with Proposition 2.8, we obtain

$$\phi''(t) = (y - x)^T \nabla^2 f(x + t(y - x))(y - x) \geq 0.$$

Hence ϕ is convex on $(0, 1)$. Moreover, ϕ is continuous on $[0, 1]$. Putting together the pieces, we find that ϕ is convex on $[0, 1]$. These considerations imply that f is convex on $\text{int}(\text{dom}(f))$.

It remains to show that f is also convex on $\text{dom}(f)$. Since f is continuous on $\text{dom}(f)$, the set $\text{dom}(f)$ is closed. Combined with Proposition 1.14, we find that each $x \in \text{dom}(f)$ is the limit point of some sequence (x^k) contained in $\text{int}(\text{dom}(f))$. Let x, $y \in \text{dom}(f)$ and let $\lambda \in [0, 1]$. There exist sequences (x^k) and (y^k) contained in $\text{int}(\text{dom}(f))$ with $x^k \to x$ and $y^k \to y$ as $k \to \infty$. For each $k \in \mathbb{N}$, we have

$$f(\lambda x^k + (1 - \lambda) y^k) \leq \lambda f(x^k) + (1 - \lambda) f(y^k).$$

Taking limits as $k \to \infty$ and using the continuity of f over $\text{dom}(f)$, we find that $f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y)$. Hence f is convex on $\text{dom}(f)$. $\square$

Theorem 2.9 can be used to establish the convexity of twice differentiable functions.

Example 2.10 (Convexity of log-sum-exp function)**.** Let us show that the log-sum-exp function $f(x) = \ln(\sum_{i=1}^n \exp(x_i))$ is convex on $\mathbb{R}^n$.

Let $d \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$. We compute

$$\begin{aligned}
\nabla f(x)^T d &= \frac{\sum_{i=1}^n \exp(x_i) d_i}{\sum_{i=1}^n \exp(x_i)}, \\
d^T \nabla^2 f(x) d &= -\frac{\left(\sum_{i=1}^n \exp(x_i) d_i\right)^2}{\left(\sum_{i=1}^n \exp(x_i)\right)^2} + \frac{\sum_{i=1}^n \exp(x_i) d_i^2}{\sum_{i=1}^n \exp(x_i)}.
\end{aligned}$$

Let us define $\lambda_i = \exp(x_i) / \sum_{i=1}^n \exp(x_i) > 0$, and $\phi(t) = t^2$. Using the fact that $\sum_{i=1}^n \lambda_i = 1$ and that ϕ is convex, Jensen's inequality implies

$$\sum_{i=1}^n \lambda_i \phi(d_i) - \phi\bigg(\sum_{i=1}^n \lambda_i d_i\bigg) \geq 0.$$

Moreover, we have

$$d^T\nabla^2 f(x)d = \sum_{i=1}^n \lambda_i d_i^2 - \left(\sum_{i=1}^n \lambda_i d_i\right)^2 = \sum_{i=1}^n \lambda_i \phi(d_i) - \phi\left(\sum_{i=1}^n \lambda_i d_i\right) \geq 0.$$

Hence f is convex.

We can establish the convexity of f using a convexity-preserving operation discussed in Section 2.3. It can be shown that $\ln(s) = \min_{z\in\mathbb{R}} [s\exp(z)-z-1]$ for any $s > 0$. Hence $\ln(\sum_{i=1}^n \exp(x_i)) = \min_{z\in\mathbb{R}} [\sum_{i=1}^n \exp(z)\exp(x_i) - z - 1]$. The objective function in the latter relation is convex in (x, z). It remains to use the convexity-preserving operation "partial minimization."

An extremely important property of convex functions is the validity of the gradient inequality.

Theorem 2.11. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function, $x \in \mathbb{R}^n$ be an interior point of the domain of f, and let X be a convex set with $x \in X$. Suppose that f is convex on X and differentiable at x. Then*

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) \quad \textit{for all} \quad y \in X. \tag{2.5}$$

Proof. Let $y \in X$. If $y \notin \mathrm{dom}(f)$ or $y = x$, then (2.5) holds true. Now let $y \neq x$ and $y \in \mathrm{dom}(f)$. We have for all $\lambda \in (0, 1)$,

$$f(x + \lambda(y - x)) = f(\lambda y + (1 - \lambda)x) \leq \lambda f(y) + (1 - \lambda)f(x).$$

Hence for all $\lambda \in (0, 1)$,

$$f(\lambda y + (1 - \lambda)x) - f(x) \leq \lambda f(y) - \lambda f(x).$$

Dividing by $\lambda \in (0, 1)$, we obtain

$$\frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \leq f(y) - f(x).$$

Taking limits as $\lambda \to 0$ and using the fact that f is differentiable at x, we obtain (2.5). □

We refer to the inequality (2.5) as *gradient inequality.* Under additional assumptions on f the gradient inequality is equivalent to the convexity of f.

Theorem 2.12. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function, let $X \subset \mathbb{R}^n$ be a convex set with nonempty interior, and let f be continuous on X and differentiable on* $\mathrm{int}(X)$. *Then f is convex on X if and only if the gradient inequality* (2.5) *is valid for every $x \in \mathrm{int}(X)$ and $y \in X$.*

Proof. The implication "⇒" is a consequence of Theorem 2.11. To establish the reverse implication, let us first show that f is convex on the interior of X. Let $\lambda \in (0, 1)$ and let x, $y \in \operatorname{int}(X)$. Let us define $z = y + \lambda(x - y)$. We have $z = \lambda x + (1-\lambda)y \in \operatorname{int}(X)$ (see Proposition 1.14). Using the gradient inequality (2.5), we obtain

$$\begin{aligned}\lambda f(x) + (1-\lambda)f(y) - f(z) &= \lambda\big(f(x) - f(z)\big) + (1-\lambda)\big(f(y) - f(z)\big)\\ &\geq \lambda \nabla f(z)^T(x - z) + (1-\lambda)\nabla f(z)^T(y - z) = 0.\end{aligned}$$

Hence f is convex on the interior of X. To establish the convexity on X, similar arguments as in the proof of Theorem 2.9 can be used. □

2.5 Lipschitz Continuity of Convex Functions*

Convex functions possess nice local properties.

Theorem 2.13. *Let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function and let X be a closed, bounded set contained in the relative interior of the domain $\operatorname{dom}(f)$ of f. Then f is Lipschitz continuous on X, that is, there exists a constant L (the Lipschitz constant of f on X) such that*

$$|f(x) - f(y)| \leq L\|x - y\|_2 \quad \textit{for all} \quad x, y \in X.$$

In particular, f is bounded on X.

Proof. See Theorem 2.4.1 in [1]. □

All three assumptions on X (1) closedness, (2) boundedness, and (3) $X \subset \operatorname{rint}(\operatorname{dom}(f))$ made in Theorem 2.13 are essential for the assertion of Theorem 2.13 to be true. We illustrate this using the following examples.

- $f(x) = 1/x$ with $\operatorname{dom}(f) = (0, +\infty)$. We consider $X = (0, 1]$. The set X is bounded and we have $X \in \operatorname{rint}(\operatorname{dom}(f))$. However, f is neither bounded nor Lipschitz continuous on X.
- $f(x) = x^2$ with $\operatorname{dom}(f) = \mathbb{R}$. We consider $X = \mathbb{R}$. The set X is closed and $X = \operatorname{int}(\operatorname{dom}(f))$. However, f is neither bounded nor Lipschitz continuous on $\mathbb{R}$.
- $f(x) = -\sqrt{x}$ with $\operatorname{dom}(f) = [0, +\infty)$. We consider $X = [0, 1]$. The set X is bounded and X is closed. However, X is not contained in the relative interior of $[0, +\infty)$. The function f fails to be Lipschitz continuous on X.

2.6 Minima of Convex Functions

We show that local minimia of convex functions are global minima.

Let $X \subset \mathbb{R}^n$ be a nonempty set and let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a function. We call a point $x^* \in \mathbb{R}^n$ a *(global) minimizer* of f over X if $x^* \in \operatorname{dom}(f) \cap X$ and $f(x^*) \leq f(x)$ for all $x \in X$. We call a point $x^* \in \mathbb{R}^n$ is a *local minimizer* of f over X if $x^* \in \operatorname{dom}(f) \cap X$ and there exists $r > 0$ such that $f(x^*) \leq f(x)$ for all $x \in X$ with $\|x - x^*\|_2 \leq r$.

Proposition 2.14. *Let $X \subset \mathbb{R}^n$ be a convex set, let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a convex function, and let $x^* \in \mathrm{dom}(f) \cap X$ be a local minimizer of f over X. Then x^* is a global minimizer of f over X.*

Proof. Since x^* is a local minimizer of f over X, there exists $r > 0$ such that $f(x^*) \leq f(y)$ for all $y \in X$ with $\|y - x^*\| \leq r$. Let $x \in X$ with $x \neq x^*$. We must show that $f(x^*) \leq f(x)$. If $f(x) = +\infty$, then we have nothing to show. Now let $f(x)$ be finite. We define

$$t = \min\left\{1, \frac{r}{\|x - x^*\|_2}\right\} \quad \text{and} \quad y = tx + (1-t)x^*.$$

We have $\|y - x^*\|_2 = \|tx - tx^*\|_2 = r$. Hence $f(x^*) \leq f(y)$. Since $f(x^*) \leq f(y)$ and f is convex, we obtain

$$f(x^*) \leq f(y) \leq tf(x) + (1-t)f(x^*).$$

Hence $f(x^*) \leq f(x)$. □

We present an alternative proof of Proposition 2.14. This proof is shorter than the above one, at the expense of being not constructive.

Proof of Proposition 2.14. Suppose that x^* is not a local minimizer of f over X. Then there exists a point $\bar{x} \in X$ such that $f(\bar{x}) < f(x^*)$. For all $t \in (0,1)$, we have $f(x^* + t(x - x^*)) \leq tf(x^*) + (1-t)f(\bar{x}) < f(x^*)$. This contradicts local optimality of x^*. □

For each function, its lower level set is convex.

Lemma 2.15. *If f is a convex function on $\mathbb{R}^n$, then for each $\alpha \in \mathbb{R}$, the level set $\{x \in \mathbb{R}^n \colon f(x) \leq \alpha\}$ is convex.*

Proof. Let x and $y \in \mathbb{R}^n$ with $f(x) \leq \alpha$ and $f(y) \leq \alpha$. For $\lambda \in [0,1]$, we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \alpha.$$

□

As a consequence of the fact that lower level sets of convex functions are convex, we obtain the convexity of the set of minimizers of a convex function f.

Proposition 2.16. *If f is a convex function on $\mathbb{R}^n$, then the set of global minimizers of f is convex.*

Proof. If f has no global minimizer, then the set of global minimizers is empty and hence convex. If x^* is a global minimizer, then we obtain the convexity of the set of minimizers from Lemma 2.15 with $\alpha = f(x^*)$. □

Our next goal is to analyze uniqueness of minimizers of a convex function. We say that a function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is *strictly convex* if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

for all $x \neq y$ and $\lambda \in (0,1)$.

Proposition 2.17. *If f is a strictly convex function, then f has at most one minimizer.*

Proof. Let x^* and y^* be minimizers of f with $x^* \neq y^*$. Since f is strictly convex, we have

$$f((1/2)x^* + (1/2)y^*) < (1/2)f(x^*) + (1/2)f(y^*) = f(x^*).$$

This is impossible as x^* is a minimizer of f. □

We now state optimality conditions for convex minimization problems.

Theorem 2.18. *Let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function, let x^* be an interior point of* $\mathrm{dom}(f)$*, and let $X \subset \mathrm{dom}(f)$ be a convex set. Suppose that f be differentiable at x^* and that f is convex on X. Then x^* is a minimizer of f over X if and only if $x^* \in X$ and*

$$\nabla f(x^*)^T (x - x^*) \geq 0 \quad \textit{for all} \quad x \in X. \tag{2.6}$$

Proof. "⇒" Let $x \in X$. Hence $x \in \mathrm{dom}(f)$. Then for all sufficiently small $\lambda \in (0,1)$,

$$0 \leq \frac{f(x^* + \lambda(x - x^*)) - f(x^*)}{\lambda}.$$

Taking limits as $\lambda \to 0$, we obtain (2.6) for all $x \in \mathrm{dom}(f)$.
"⇐" The gradient inequality (2.5) implies the assertion. □

Remark 2.19. The conditions $x^* \in X$ and the "variational inequality" (2.6) are called *first-order optimality conditions.* Actually, the proof of Theorem 2.18 shows that these first-order optimality conditions are valid without convexity of f.

We discuss an equivalent form of the optimality condition (2.6). Let the hypotheses of Theorem 2.18 hold true and consider a point $\bar{x} \in X$. We define the *radial cone* (or *tangent cone*) of X at $\bar{x}$ by

$$T_X(\bar{x}) = \{d \in \mathbb{R}^n : \bar{x} + td \in X \quad \text{for all sufficiently small } t > 0\}.$$

Since X is convex, the set $T_X(\bar{x})$ is a cone consisting of all vectors of the form $t(x - \bar{x})$, where $x \in X$ and $t \geq 0$. Using the radial cone, we can rewrite (2.6) as

$$\nabla f(x^*)^T d \geq 0 \quad \text{for all} \quad d \in T_X(x^*). \tag{2.7}$$

The optimality conditions (2.6) and (2.7) are equivalent, as X is a convex set. Moreover, (2.7) is equivalent to

$$\nabla f(x^*) \in N_X(x^*), \tag{2.8}$$

where $N_X(x^*)$ is the normal cone of X at x^*. For $\bar{x} \in X$, the normal cone of X at $\bar{x}$ is defined by

$$N_X(\bar{x}) = \{g \in \mathbb{R}^n \colon g^T d \geq 0 \quad \text{for all} \quad d \in T_X(\bar{x})\}. \tag{2.9}$$

Let us provide examples.

Example 2.20. We consider the feasible set

$$X = \{x \in \mathbb{R}^n \colon a_i^T x \leq b_i, \quad i = 1, \dots, m\}.$$

The radial cone of X at $\bar{x} \in X$ is given by

$$T_X(\bar{x}) = \{d \in \mathbb{R}^n \colon a_i^T d \leq 0, \quad \text{for all} \quad i \in \mathcal{A}(\bar{x})\},$$

where $\mathcal{A}(\bar{x})$ is the set of active constraints at $\bar{x}$:

$$\mathcal{A}(\bar{x}) = \{i \in \{1, \dots, m\} \colon a_i^T \bar{x} = b_i\}.$$

Let us verify this identity for the tangent cone. Let $d \in \mathbb{R}^n$ and $t > 0$. We have

$$a_i^T(\bar{x} + td) = a_i^T \bar{x} + t a_i^T d = b_i + t a_i^T d, \quad \text{for all} \quad i \in \mathcal{A}(\bar{x}).$$

"$\subset$" Let $\bar{x} + td \in X$ for all sufficiently small $t > 0$. Fix a sufficiently small $t > 0$ and $i \in \mathcal{A}(\bar{x})$. Using $b_i \geq b_i + t a_i^T d$, we obtain $a_i^T d \leq 0$ for $i \in \mathcal{A}(\bar{x})$.

"$\supset$" Now let $d \in \mathbb{R}^n$ with $a_i^T d \leq 0$ for $i \in \mathcal{A}(\bar{x})$. Fix $t > 0$ and $i \in \mathcal{A}(\bar{x})$. Using $a_i^T d \leq 0$, we obtain $a_i^T(\bar{x} + td) \leq b_i$.

Example 2.21. If $X \subset \mathbb{R}^n$ is a set and $\bar{x} \in \operatorname{int}(X)$, then $T_X(\bar{x}) = \mathbb{R}^n$ and $N_X(\bar{x}) = \{0\}$. Let us verify these assertions. Since $\bar{x}$ is an interior point of X, there exists a positive radius r such that the set $\{y \in \mathbb{R}^n \colon \|x - y\|_2 < r\}$ is contained in X. Therefore each $d \in \mathbb{R}^n$ is contained in $T_X(\bar{x})$. Hence $T_X(\bar{x}) = \mathbb{R}^n$. Using the definition of the normal cone provided in (2.9), we find that $N_X(\bar{x}) = \{0\}$.

Therefore, if $x^* \in \operatorname{int}(X)$, then (2.6) and (2.8) become the Fermat condition $\nabla f(x^*) = 0$.

Example 2.22. Let $x^* \in \operatorname{rint}(X)$. Since $\operatorname{Aff}(X)$ is an affine subspace of $\mathbb{R}^n$, we have $\operatorname{Aff}(X) = x^* + L$, where L is a linear subspace in $\mathbb{R}^n$. We obtain $T_X(x^*) = L$ and hence $N_X(x^*) = L^\perp$, where $L^\perp = \{d \in \mathbb{R}^n \colon d^T h = 0 \quad \text{for all} \quad h \in L\}$ is the orthogonal complement of L. Therefore (2.8) becomes

$$\nabla f(x^*) \quad \text{is orthogonal to} \quad L.$$

2.7 Optimization over Polyhedral Sets

Below we discuss the optimality conditions of minimizing a convex function over a polyhedral set. To derive the optimality conditions, we use the homogeneous Farkas lemma.

We consider the homogeneous linear inequality

$$a^T x \geq 0 \tag{2.10}$$

and the finite system of the inequalities

$$a_i^T x \geq 0, \quad i = 1, \ldots, m. \tag{2.11}$$

Here a, $a_1, \ldots, a_m \in \mathbb{R}^n$ are vectors. If a is a conic combination of $a_1, \ldots, a_m$ with

$$a = \sum_{i=1}^{m} \lambda_i a_i \quad \text{with} \quad \lambda_i \geq 0, \tag{2.12}$$

then (2.10) is a consequence of (2.11). In other words, if a is a conic combination of $a_1, \ldots, a_m$ and $x \in \mathbb{R}^n$ satisfies (2.11), then x satisfies (2.10). Let us verify this claim. Using (2.12), we obtain

$$a^T x = \sum_{i=1}^{m} \lambda_i a_i^T x \quad \text{for all} \quad x \in \mathbb{R}^n.$$

Hence $a^T x \geq 0$ if $a_i^T x \geq 0$ for $i = 1, \ldots, m$.

Lemma 2.23 (Homogeneous Farkas lemma)**.** *The inequality* (2.10) *a consequence of* (2.11) if and only if *a is a conic combination of* $a_1, \ldots, a_m$.

Before we establish Lemma 2.23, we briefly mention a geometric interpretation of the Farkas lemma. If $a \notin \mathrm{Cone}(\{a_1, \ldots, a_m\})$, then $\{a\}$ and $\mathrm{Cone}(\{a_1, \ldots, a_m\})$ can be separated by a hyperplane. If x solves (2.11) and $a^T x < 0$, then $\mathrm{Cone}(\{a_1, \ldots, a_m\})$ is contained in the half space $\{c \in \mathbb{R}^n \colon c^T x \geq 0\}$ and a is contained in the open half space $\{c \in \mathbb{R}^n \colon c^T x < 0\}$. See also Exercise 1.19.

Proof of Lemma 2.23. "$\Leftarrow$" We have already established this direction of the lemma.

"$\Rightarrow$" We show that if a is not a conic combination of $a_1, \ldots, a_m$, then there exists $d \in \mathbb{R}^n$ such that $a^T d < 0$ and $a_i^T d \geq 0$, $i = 1, \ldots, m$. The set $K \coloneqq \mathrm{Cone}(\{a_1, \ldots, a_m\})$ is polyhedrally representable:

$$\mathrm{Cone}(\{a_1, \ldots, a_m\}) = \left\{ x \in \mathbb{R}^n \colon \exists \lambda \in \mathbb{R}^m_+ \quad \text{with} \quad x = \sum_{i=1}^{n} \lambda_i a_i \right\}.$$

Theorem 1.30 ensures that K is polyhedral. Hence there exist $L \in \mathbb{N}$, $d_\ell \in \mathbb{R}^n$, and $c_\ell \in \mathbb{R}$ with

$$K = \{x \in \mathbb{R}^n \colon d_\ell^T x \geq c_\ell, \quad \ell = 1, \ldots, L\}.$$

Since $0 \in K$, we have $c_\ell \le 0$ for $\ell = 1, \ldots, L$. Since $\lambda a_i \in K$ for all $\lambda > 0$, we have $\lambda d_\ell^T a_i \ge c_\ell$ for all $\lambda > 0$. Hence $d_\ell^T a_i \ge 0$ for $i = 1, \ldots, m$ and $\ell = 1, \ldots, L$. If $a \notin K$, then we can find an index $\ell \in \{1, \ldots, L\}$ with $d_\ell^T a < c_\ell$. Since $c_\ell \le 0$, we have $d_\ell^T a < 0$. Choosing $d = d_\ell$, we find that $a^T d < 0$ and $a_i^T d \ge 0$, $i = 1, \ldots, m$. □

We derive necessary and sufficient optimality conditions for the minimization of a convex function over the polyhedral set

$$X = \{x \in \mathbb{R}^n \colon a_i^T x \le b_i, \quad i = 1, \ldots, m\}. \tag{2.13}$$

Let the hypotheses of Theorem 2.18 hold true. As discussed in Example 2.20, the radial cone of X at $\bar{x}$ is given by

$$T_X(\bar{x}) = \{d \in \mathbb{R}^n \colon a_i^T d \le 0, \quad \text{for all} \quad i \in \mathcal{A}(\bar{x})\},$$

where $\mathcal{A}(\bar{x})$ is the set of active constraints at $\bar{x}$

$$\mathcal{A}(\bar{x}) = \{i \in \{1, \ldots, m\} \colon a_i^T \bar{x} = b_i\}.$$

Using the Farkas lemma, Lemma 2.23, we find that the normal cone of X at $\bar{x}$ is given by

$$\begin{aligned} N_X(\bar{x}) &= \{g \in \mathbb{R}^n \colon a_i^T d \le 0,\ i \in \mathcal{A}(\bar{x}),\ d \in \mathbb{R}^n \quad \text{implies} \quad g^T d \ge 0\} \\ &= \Big\{g \in \mathbb{R}^n \colon \exists \lambda_i \ge 0 \quad \text{with} \quad g = -\sum_{i \in \mathcal{A}(\bar{x})} \lambda_i a_i\Big\}. \end{aligned}$$

Hence the optimality condition (2.8) is equivalent to: there exist $\lambda_i^* \ge 0$ for $i \in \mathcal{A}(x^*)$ with $\nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* a_i = 0$.

Let the hypotheses of Theorem 2.18 hold true. Combining our derivations, we find that x^* is a minimizer of the convex, differentiable function f over the polyhedral set $X = \{x \in \mathbb{R}^n \colon a_i^T x \le b_i, \quad i = 1, \ldots, m\}$ defined in (2.13) if and only if there exists $\lambda^* \in \mathbb{R}^m$ with

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* a_i &= 0, \\ \lambda_i^* &\ge 0, \quad i = 1, \ldots, m, \\ \lambda_i^*(a_i^T x^* - b_i) &= 0, \quad i = 1, \ldots, m, \\ a_i^T x^* - b_i &\le 0, \quad i = 1, \ldots, m. \end{aligned} \tag{2.14}$$

Under the hypotheses of Theorem 2.18 on f, the conditions (2.14) provide necessary and sufficient optimality conditions for the optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad a_i^T x \le b_i, \quad i = 1, \ldots, m.$$

Let us now consider the optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad a_i^T x \le b_i, \quad i=1,\dots,m, \quad c_j^T x = d_j, \quad j=1,\dots,p, \tag{2.15}$$

where $c_j \in \mathbb{R}^n$ and $d_j \in \mathbb{R}$. Reformulating the equality constraints $c_j^T x = d_j$ as the inequality constraints $c_j^T x \le d_j$ and $-c_j^T x \le -d_j$ and applying (2.14), we find that x^* is a solution to (2.15) if and only if there exists $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ with

$$\begin{aligned}\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* a_i + \sum_{j=1}^p \mu_j^* c_j = 0,&\\ \lambda_i^* \ge 0, &\quad i=1,\dots,m,\\ \lambda_i^*(a_i^T x^* - b_i) = 0, &\quad i=1,\dots,m,\\ a_i^T x^* - b_i \le 0, &\quad i=1,\dots,m,\\ c_j^T x^* - d_j = 0, &\quad j=1,\dots,p,\end{aligned}$$

provided that f is convex and differentiable at x^*.

Example 2.24. Let us solve the problem

$$\min_{x\in\mathbb{R}^n} c^T x + \sum_{i=1}^n x_i \ln(x_i) \quad \text{s.t.} \quad x \ge 0, \quad \sum_{i=1}^n x_i = 1. \tag{2.16}$$

For $x_i = 0$, we define $x_i \ln(x_i) = 0$. The feasible set $X \coloneqq \{x \in \mathbb{R}^n \colon x \ge 0, \ \sum_{i=1}^n x_i = 1\}$ is convex and the objective function $f(x) \coloneqq c^T x + \sum_{i=1}^n x_i \ln(x_i)$ is convex and continuous on X and differentiable on the set $\{x \in \mathbb{R}^n \colon x > 0\}$.

Let us compute the gradient of f. For $x \in \mathbb{R}^n$ with $x > 0$, we have

$$\nabla f(x) = c + \begin{bmatrix} \ln(x_1) + 1 \\ \vdots \\ \ln(x_n) + 1 \end{bmatrix}$$

The constraint $\sum_{i=1}^n x_i = 1$ can be written as two inequality constraints $a_1^T x \le 1$ and $a_2^T x \le -1$ with $a_1 = (1,\dots,1)$ and $a_2 = (-1,\dots,-1)$.

Let us *assume* that $x^* \in \mathbb{R}^n$ with $x^* > 0$ is a solution to (2.16). Then we must have $\sum_{i=1}^n x_i^* = 1$ and for some $\lambda^* \in \mathbb{R}$,

$$\nabla_x \left[f(x) + \lambda^* \left(\sum_{i=1}^n x_i - 1 \right) \right]_{x=x^*} = 0.$$

These n equations are equivalent to $\ln(x_i^*) = -c_i - \lambda - 1$ for $i = 1,\dots,n$. Hence $x_i^* = \exp(-1-\lambda)\exp(-c_i)$. Since we must have $\sum_{i=1}^n x_i^* = 1$, we obtain

$$x_i^* = \frac{\exp(-c_i)}{\sum_{j=1}^n \exp(-c_j)}$$

This point satisfies the optimality conditions and hence is a solution to (2.16).

2.8 Subgradients

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $\bar{x} \in \operatorname{int}(\operatorname{dom}(f))$. If f is differentiable at $\bar{x}$, then the gradient inequality ensures that the affine function

$$h(x) := f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$$

satisfies

$$f(x) \geq h(x) \quad \text{for all} \quad x \in \mathbb{R}^n, \quad \text{and} \quad f(\bar{x}) = h(\bar{x}). \tag{2.17}$$

Hence h is an affine minorant and $h(\bar{x}) = f(\bar{x})$. Such an affine function may also exist in the case that f is nondifferentiable at $\bar{x} \in \operatorname{dom}(f)$ in the sense that for some $g \in \mathbb{R}^n$, the function

$$h(x) := f(\bar{x}) + g^T (x - \bar{x})$$

satisfies the conditions in (2.17).

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function and let $\bar{x} \in \operatorname{dom}(f)$. A vector $g \in \mathbb{R}^n$ such that

$$f(x) \geq f(\bar{x}) + g^T (x - \bar{x}) \quad \text{for all} \quad x \in \mathbb{R}^n \tag{2.18}$$

is called a *subgradient* of f at $\bar{x}$. The set of all subgradients of f at $\bar{x}$ is called *subdifferential* of f at $\bar{x}$ and is denoted by $\partial f(\bar{x})$. We refer to the inequality (2.18) as *subgradient inequality* of f at $\bar{x} \in \operatorname{dom}(f)$.

Example 2.25. If $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex function and differentiable at $\bar{x} \in \operatorname{int}(\operatorname{dom}(f))$, then $\nabla f(\bar{x})$ is the unique element of $\partial f(\bar{x})$, that is, $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.

The gradient inequality ensures that $\nabla f(\bar{x}) \in \partial f(\bar{x})$. Let $g \in \partial f(\bar{x})$. We show that $g = \nabla f(\bar{x})$. Fix $h \in \mathbb{R}^n$ and $t \in \mathbb{R}^n$. Using the subgradient inequality, we obtain $f(\bar{x} + th) \geq f(\bar{x}) + tg^T h$. Rearranging terms and dividing by $t > 0$, we have $t^{-1}(f(\bar{x} + th) - f(\bar{x})) \geq g^T h$. Taking limits as $t \to 0^+$, we find that $\nabla f(\bar{x})^T h \geq g^T h$. Hence $0 \geq (g - \nabla f(\bar{x}))h$. Since this inequality is valid for all $h \in \mathbb{R}^n$, we obtain $g = \nabla f(\bar{x})$.

Example 2.26. We consider $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$. If $\bar{x} \neq 0$, then f is differentiable at $\bar{x}$ and hence $\partial f(\bar{x}) = \{f'(\bar{x})\}$. If $\bar{x} = 0$, then g is a subgradient of f at 0 if and only if

$$|x| \geq 0 + gx = gx \quad \text{for all} \quad x \in \mathbb{R}.$$

We obtain $\partial f(0) = [-1, 1]$.

Using the separation theorem, we can show the existence of subgradients.

Proposition 2.27. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function. If $x \in \operatorname{rint}(\operatorname{dom}(f))$, then $\partial f(x)$ is nonempty.*

Proof. We present a proof assuming $x \in \operatorname{int}(X)$. The point $(x, f(x)) \in \mathbb{R}^{n+1}$ is not a point from the relative interior of the convex set $\operatorname{epi}(f) = \{(x,t) \in \mathbb{R}^{n+1} : f(x) \le t\}$, since $(x, f(x) - r)$ is not contained in $\operatorname{epi}(f)$ for all $r > 0$. Therefore, the relative interiors of $\operatorname{epi}(f)$ and $\{(x, f(x))\}$ are disjoint. The separation theorem, Theorem 1.34, ensures the existence of a vector $(a, \alpha) \in \mathbb{R}^{n+1} \setminus \{0\}$ with

$$(a,\alpha)^T(y,t) \le (a,\alpha)^T(x, f(x)) \quad \text{for all} \quad (y,t) \in \operatorname{epi}(f). \tag{2.19}$$

This is equivalent to

$$a^T y + \alpha t \le a^T x + \alpha f(x) \quad \text{for all} \quad (y,t) \in \operatorname{epi}(f).$$

Since $(y, f(x)+1) \in \operatorname{epi}(f)$, we have $\alpha \le 0$. Suppose that $\alpha = 0$. Then we have

$$a^T y \le a^T x.$$

Since $x \in \operatorname{int}(X)$, there exists $t > 0$ with $y := x + ta \in X$. Since $(y, f(y)) \in \operatorname{epi}(f)$, we obtain

$$a^T y = a^T x + t a^T a = a^T x + t\|a\|_2^2 \ge a^T y + t\|a\|_2^2 > a^T y.$$

This is a contradiction. Hence $\alpha < 0$.

Multiplying (2.19) by $1/|\alpha|$ and defining $g = a/|\alpha|$, we obtain

$$g^T y - z \le g^T x - f(x) \quad \text{for all} \quad (y,z) \in \operatorname{epi}(f).$$

Let us now choose $z = f(y)$. We obtain

$$f(y) \ge f(x) + g^T(y - x) \quad \text{for all} \quad y \in \mathbb{R}^n.$$

□

The following example shows that a convex function f may lack subgradients at $x \in \operatorname{dom}(f) \setminus \operatorname{rint}(f)$.

Example 2.28. We define $f\colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ by $f(x) := \sqrt{x}$ if $x \ge 0$ and $f(x) := +\infty$ otherwise. The function f is convex (because, for example, f is continuous on $\operatorname{dom}(f)$, and twice differentiable on $(0,+\infty)$ with positive second derivative). We have $0 \notin \operatorname{rint}(\operatorname{dom}(f)) = (0, = \infty)$ and $\partial f(0) = \emptyset$. The slope of f at 0 fails to be defined, resulting in $\partial f(0) = \emptyset$. Let use verify this statement. Suppose that $g \in \partial f(0)$. Then $-\sqrt{x} \ge gx$ for all $x \ge 0$. Hence $-\sqrt{1/x} \ge g$ for all $x > 0$. Since $-\sqrt{1/x} \to -\infty$ as $x \to 0^+$, we cannot have $g \in \mathbb{R}$.

Below we provide some basic subgradient calculus for convex functions. Observe that many of them mimic the calculus for gradient computation.

1. Scaling: If $\lambda > 0$ and $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, then $\partial[\lambda f] = \lambda \partial f$.
2. Addition: If $X \subset \mathbb{R}^n$, $f\colon X \to \mathbb{R}$ and $g\colon X \to \mathbb{R}$ are convex, then $\partial[f+g] = \partial f + \partial g$.

3. Affine composition: If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then $g(x) := f(Ax+b)$ is convex with $\partial g(x) = A^T \partial f(Ax+b)$ provided that $Ax+b \in \mathrm{dom}(f)$.
4. Finite pointwise maximum: If $X \subset \mathbb{R}^n$ is convex and open, $f_i \colon X \to \mathbb{R}$, $i = 1, \dots, m$, are convex, and $f(x) := \max_{1 \le i \le m} f(x)$, then

$$\partial f(x) = \mathrm{Conv}(\cup_{i\colon f_i(x)=f(x)} \partial f_i(x)).$$

   This is the convex hull of the union of subdifferentials corresponding to functions f_i with $f_i(x) = f(x)$.

We conclude this section by stating some important results of subgradients without providing the proof. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex.

1. For every $x \in \mathrm{dom}(f)$, $\partial f(x)$ is closed and convex.
2. If $x \in \mathrm{rint}(\mathrm{dom}(f))$, then for every $h \in \mathbb{R}^n$, the directional derivative $f'(x;h)$ of f at x in direction h, that is,

$$f'(x;h) := \lim_{t \to 0^+} \frac{f(x+th) - f(x)}{t},$$

   exists and

$$f'(x;h) = \max_{g \in \partial f(x)} g^T h.$$

3. Let $(x^k) \subset \mathrm{dom}(f)$ with $x^k \to x \in \mathrm{dom}(f)$. Suppose that

$$f(x) \le \liminf_{k \to \infty} f(x^k).$$

   If $g^k \in \partial f(x^k)$ converges to some $g \in \mathbb{R}^n$ as $k \to \infty$, then $g \in \partial f(x)$.
4. The multi-valued mapping $x \mapsto \partial f(x)$ is locally bounded about every $\bar{x} \in \mathrm{int}(\mathrm{dom}(f))$, that is, whenever $\bar{x} \in \mathrm{int}(\mathrm{dom}(f))$, then there exist $r > 0$ and $R < \infty$ such that if

$$\|x - \bar{x}\|_2 \le r \quad \text{and} \quad g \in \partial f(x),$$

   then $\|g\|_2 \le R$.

2.9 Exercises

Exercise 2.1.
Prove Theorem 2.3.

Exercise 2.2 (Minimizing linear functions)**.**
Let $X \subset \mathbb{R}^n$ be a nonempty set, and let $c \in \mathbb{R}^n$. Show that

$$\inf_{x \in \operatorname{Conv}(X)} c^T x = \inf_{x \in X} c^T x.$$

Exercise 2.3.
Which of the following functions are convex on their indicated domains? No justifications are required.

1. $f(x) = 1$ on $\mathbb{R}$.
2. $f(x) = x$ on $\mathbb{R}$.
3. $f(x) = |x|$ on $\mathbb{R}$.
4. $f(x) = -|x|$ on $\mathbb{R}$.
5. $f(x) = -|x|$ on $\mathbb{R}_+ = \{x \in \mathbb{R} \colon x \geq 0\}$.
6. $f(x) = \exp(x)$ on $\mathbb{R}$.
7. $f(x) = \exp(x^2)$ on $\mathbb{R}$.
8. $f(x) = \exp(-x^2)$ on $\mathbb{R}$.
9. $f(x) = \exp(-x^2)$ on $\{x \in \mathbb{R} \colon x \geq 100\}$.
10. $f(x) = x \ln(x)$ on $\{x \in \mathbb{R} \colon x > 0\}$.
11. $f(x) = \sin(x)$ on $\mathbb{R}$.
12. $f(x) = -\ln(x)$ on $\{x \in \mathbb{R} \colon x > 0\}$.

Exercise 2.4 (Maximum of a convex function over a polyhedron)**.**
Show that if f is a convex function on $\mathbb{R}^n$ and $X = \operatorname{Conv}(x^1, \ldots, x^m)$ with $x^i \in \mathbb{R}^n$, then

$$\sup_{x \in X} f(x) = \max_{i=1,\ldots,m} f(x^i).$$

Hint: Use Jensen's inequality.

Exercise 2.5.
Show that the following functions are convex on the indicated domains:

1. $f(x, y) = x^2/y$ on $\{(x, y) \in \mathbb{R}^2 \colon y > 0\}$.
2. $f(x, y) = 1/(xy)$ on $\{(x, y) \in \mathbb{R}^2 \colon x > 0, y > 0\}$.

Exercise 2.6 (Products and ratios of convex functions)**.**
In general the product or ratio of two convex functions is not convex. However, there are some results that apply to functions on $\mathbb{R}$. Prove the following.

1. If f and g are convex, both nondecreasing, and positive functions on an interval, then fg is convex.
2. If f, g are concave, positive, with one nondecreasing and the other nonincreasing, then fg is concave.
3. If f is convex, nondecreasing, and positive, and g is concave, nonincreasing, and positive, then f/g is convex.

Exercise 2.7.
In Example 2.10, we have discussed two approaches to show that the log-sum-exp function $f(x) = \ln(\sum_{i=1}^n \exp(x_i))$ is a convex function on $\mathbb{R}^n$. In this exercise, we establish the convexity of the log-sum-exp function using the facts that a function is convex if and only if its epigraph is a convex set and that level sets of convex functions are convex.

1. Show that
$$\operatorname{epi}(f) = \left\{(x, t) \in \mathbb{R}^n \times \mathbb{R} \colon \sum_{i=1}^n \exp(x_i - t) \leq 1\right\}.$$
2. Deduce the convexity of the log-sum-exp function f.

Exercise 2.8.
Establish the following statements.

1. If $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$, then the function $f(x) = \ln(\sum_{i=1}^n \exp(a_i^T x + b_i))$ is convex.
2. If $A \in \mathbb{R}^{n\times n}$ is symmetric positive semidefinite, then the function $f(x) = \exp(x^T Ax)$ is convex.

3. If f is concave on $\mathbb{R}^n$ and $0 < f(x) < +\infty$ for all $x \in \mathbb{R}^n$, then $g(x) = 1/f(x)$ is convex.

4. If $p \in [1, \infty)$ and $\|\cdot\|$ is a norm on $\mathbb{R}^n$, then $\|\cdot\|^p$ is convex.

5. The function $f(x,t) := x^T x/t$ is convex over $\{(x,t) \in \mathbb{R}^n \times \mathbb{R} \colon t > 0\}$.

6. The function $f \colon \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} (1/2)\|x\|_2^2 & \text{if} \quad \|x\|_2 \leq 1, \\ \|x\|_2 - (1/2) & \text{otherwise,} \end{cases}$$

   is convex. *Hint*: Show that $f(x) = \sup_{\|y\|_2 \leq 1} x^T y - (1/2)\|y\|_2^2$.

Exercise 2.9 (Inequality of arithmetic and geometric means).
Show that if $\lambda_1, \ldots, \lambda_n$ are positive scalars with $\sum_{i=1}^n \lambda_i = 1$, then for every set of nonnegative scalars $x_1, \ldots, x_n$, we have

$$x_1^{\lambda_1} \cdot x_2^{\lambda_2} \cdots x_n^{\lambda_n} \leq \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n$$

with equality if and only if $x_1 = x_2 = \cdots = x_n$.
Hint: Show that $-\ln(x)$ is strictly convex on $(0, +\infty)$.

Exercise 2.10.
A function f defined on a convex set X is called *log-convex* on X if it takes positive values on X and if the function $\ln f$ is convex on X. Show that

1. a log-convex function on X is convex on X.

2. the sum of two log-convex functions on X is log-convex on X.

   Hint: Use the fact that the log-sum-exp function is convex (see Exercise 2.7) and use your knowledge on operations that preserve convexity.

Exercise 2.11.
Let $\phi : \mathbb{R} \to [0, \infty) \cup \{+\infty\}$ be nondecreasing. Let $a \in \mathbb{R}$ and suppose that $\phi(a)$ is finite. Show that the function

$$f(x) = \int_a^x \phi(t) dt$$

is convex.

Exercise 2.12 (Characterization of convexity using gradient monotonicity).
Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function, let $X \subset \mathbb{R}^n$ be a convex set with nonempty interior, and let f be continuous on X and differentiable on $\text{int}(X)$. Then f is convex on X if and only if

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0 \quad \text{for all} \quad x, y \in \text{int}(X). \tag{2.20}$$

Hint: To show that (2.20) implies convexity of f, use the identity

$$f(y) = f(x) + \int_0^1 \nabla f(x + t(y - x))^T (y - x) dt$$

and the characterization of convexity provided in Theorem 2.12.

Exercise 2.13 (Strong convexity).
Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function and let $X \subset \text{dom}(f)$ be a nonempty, convex set. Furthermore, let $\sigma > 0$ be a scalar. We say that f is *strongly convex over X with parameter (or coefficient) σ* if for all x, $y \in X$ and $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1 - \lambda)y) + (\sigma/2)\lambda(1 - \lambda)\|x - y\|_2^2 \leq \lambda f(x) + (1 - \lambda)f(y).$$

Note that the Euclidean norm is used in the above inequality.

Strongly convex functions are particularly "nice" functions when it comes to optimization. This problem establishes characterizations of strongly convex functions using first- and second-order derivative criteria, and implications of strong convexity. We use key properties of convex functions, such as the gradient inequality, to establish these characterizations. See Exercise 2.14 *for examples of strongly convex functions.*

1. Show that if f is strongly convex over X with coefficient σ, then f is strictly convex over X.

2. Show that f is strongly convex over X with coefficient σ if and only if the function $g(x) := f(x) - (\sigma/2)\|x\|_2^2$ is convex over X.

   This characterization of strong convexity is **extremely** useful.

3. Suppose that $\text{int}(X)$, the interior of X, is nonempty and that f is continuously differentiable on $\text{int}(X)$. Show that the following statements are equivalent:

   a) f is strongly convex over X with parameter σ.

   b) We have
   $$f(y) \geq f(x) + \nabla f(x)^T (y - x) + (\sigma/2)\|x - y\|_2^2 \quad \text{for all} \quad x, y \in \text{int}(X).$$

   c) We have
   $$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \sigma\|x - y\|_2^2 \quad \text{for all} \quad x, y \in \text{int}(X).$$

*Hints: To establish "*a) *implies* b)*", use the second part and the gradient inequality. Exercise* 2.12 *may also be helpful.*

4. If in addition to the conditions in part three, f is twice continuously differentiable on $\mathrm{int}(X)$, then the conditions a)–c) are equivalent to $\nabla^2 f(x) - \sigma I$ is positive semidefinite for each $x \in \mathrm{int}(X)$.

Exercise 2.14 (Examples of strongly convex functions)**.**
This problem highlights an important strongly convex function. Moreover it establishes the fact that sums of convex and strongly convex functions are strongly convex.

1. Show that the function $(1/2)\|\cdot\|_2^2$ is strongly convex over $\mathbb{R}^n$ with parameter 1.

2. Show that if f is strongly convex over a nonempty convex set $X \subset \mathbb{R}^n$ with parameter $\sigma > 0$, and g is convex over X, then $f + g$ is strongly convex over X with parameter σ.

Exercise 2.15.
Provide counterexamples for the following statements.

1. If a function is strictly convex, then it is strongly convex.

2. If a function is strictly convex, then it has a minimizer.

Exercise 2.16 (Minimization of strongly convex functions)**.**
Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function that is continuous and strongly convex over a nonempty, closed, convex set X with parameter $\sigma > 0$.

Show that there exists a unique point $x^* \in X$ that minimizes f over X and that

$$f(x) \geq f(x^*) + (\sigma/2)\|x - x^*\|_2^2 \quad \text{for all} \quad x \in X.$$

Inequalities of this type are often referred to as quadratic growth conditions.

This problem demonstrates that every continuous, strongly convex function over a nonempty closed convex set has a unique minimizer and exhibits quadratic growth around it.

You may establish your own proof or solve the following subproblems.

1. Let us define $g(x) = f(x) - (\sigma/2)\|x\|_2^2$. Show that there exists $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that

$$g(x) \geq a^T x + b \quad \text{for all} \quad x \in X.$$

2. Let $x_0 \in X$. Deduce that the level set $\{x \in X : f(x) \leq f(x_0)\}$ is bounded.

3. Using the fact that a continuous function on a nonempty, closed, bounded set in $\mathbb{R}^n$ has a minimizer, show that f has a minimizer over X.

4. Complete the proof.

Exercise 2.17 (Convexity under composition, see [3, Exercise 1.3]).
Let C be a nonempty convex subset of $\mathbb{R}^n$. Let also $f = (f_1, \dots, f_m)$, where $f_i : C \to \mathbb{R}$, $i = 1, \dots, m$, are convex functions. Let $g : \mathbb{R}^m \to \mathbb{R}$ be convex and monotonically non-decreasing, in the sense that $u \leq v$ implies $g(u) \leq g(v)$. Show that the function h defined by $h(x) = g(f(x))$ is convex over C. Show that if in addition $m = 1$, g is monotonically increasing, and f is strictly convex, then h is strictly convex.

Exercise 2.18 (Optimality conditions for linear programs).
We consider the linear program

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax \leq b,$$

where $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$. The system of inequalities $Ax \leq b$ can be written as $a_i^T x \leq b_i$, $i = 1, \dots, m$. State the optimality conditions derived in (2.14) for the linear program.

Exercise 2.19 (Midpoint convexity*).
A function $f : \mathbb{R} \to \mathbb{R}$ is called midpoint convex if for all x, $y \in \mathbb{R}$, $f((x + y)/2) \leq (1/2)f(x) + (1/2)f(y)$. Here, the point $(x + y)/2$ is called the midpoint of x and y. Show that a *continuous* midpoint convex function defined on $\mathbb{R}$ is convex.

Hint: The statement can be established using a proof by contradiction or using the fact that dyadic rational numbers are dense in $\mathbb{R}$.

Exercise 2.20 (Inhomogeneous Farkas Lemma).
Let $\bar{x}$ be a solution to $a_i^T x \leq b_i$, $i = 1, \dots, m$, where $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

The inequality

$$a^T x \leq b$$

is a consequence of

$$a_i^T x \leq b_i, \quad i = 1, \dots, m,$$

if and only if there exist scalars $\lambda_i \geq 0$, $i = 1, \ldots, m$, such that

$$a = \sum_{i=1}^{m} \lambda_i a_i, \quad b \geq \sum_{i=1}^{m} \lambda_i b_i. \tag{2.21}$$

This statement is known as the inhomogeneous Farkas Lemma.

Establish the inhomogeneous Farkas Lemma. You may either provide your own proof or solve the following subproblems.

1. Show that if there exist nonnegative scalers λ_i such that (2.21) holds, then $a^T x \leq b$ for all $x \in \mathbb{R}^n$ with $a_i^T x \leq b_i$, $i = 1, \ldots, m$.
2. Suppose that $a^T x \leq b$ for all $x \in \mathbb{R}^n$ with $a_i^T x \leq b_i$, $i = 1, \ldots, m$. Consider the linear program

$$\min_{x \in \mathbb{R}^n} b - a^T x \quad \text{s.t.} \quad a_i^T x \leq b_i, \quad i = 1, \ldots, m.$$

   a) Show that the linear program has a solution.
   b) Use the optimality conditions (2.14) to establish the existence of nonnegative scalers λ_i such that (2.21) holds.

Exercise 2.21 (Fermat's rule)**.**
Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex. Let $x^* \in \operatorname{dom}(f)$. Show that x^* is a minimizer of f if and only if $0 \in \partial f(x^*)$.

Exercise 2.22.
Let $X \subset \mathbb{R}^n$ be a nonempty set, and define $I_X \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ by $I_X(x) \coloneqq 0$ if $x \in X$ and $I_X(x) \coloneqq +\infty$ otherwise. Show that X is convex if and only if I_X is convex.

Exercise 2.23 (Quasi-convex functions)**.**
Let $X \subset \mathbb{R}^n$ be nonempty and convex. A function $f \colon X \to \mathbb{R}$ is called *quasi-convex* if for all x, $y \in X$ and $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda) y) \leq \max\{f(x), f(y)\}.$$

Let $X \subset \mathbb{R}^n$ be nonempty and convex. Show that $f \colon X \to \mathbb{R}$ is quasi-convex if and only if $\{x \in X \colon f(x) \leq \alpha\}$ is convex for each $\alpha \in \mathbb{R}$.

Exercise 2.24 (Proximal operator)**.**
For a convex, proper, lower semicontinuous function $g \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, we define its prox-mapping by

$$\operatorname{prox}_g(x) \coloneqq \operatorname{argmin}_{y \in \mathbb{R}^n} (1/2)\|y - x\|_2^2 + g(y).$$

Show that if $g \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, proper, lower semicontinuous function, then prox_g is well-defined. (prox_g is well-defined if for each $x \in \mathbb{R}^n$, the minimization problem defining $\operatorname{prox}_g(x)$ has a unique solution.)

Exercise 2.25 (Optimality conditions for composite convex optimization I)**.**
Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable, and let $\psi \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex. Let $x^* \in \operatorname{dom}(\psi)$.

Show that x^* is a solution to

$$\min_{x \in \mathbb{R}^n} f(x) + \psi(x)$$

if and only if

$$\nabla f(x^*)^T (x - x^*) + \psi(x) - \psi(x^*) \geq 0 \quad \text{for all} \quad x \in \mathbb{R}^n$$

if and only if

$$-\nabla f(x^*) \in \partial \psi(x^*).$$

Note: This statement generalizes Theorem 2.18.

Exercise 2.26 (Optimality conditions for composite convex optimization II)**.**
Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable, and let $\psi \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex. Let $x^* \in \operatorname{dom}(\psi)$.

Show that x^* is a solution to

$$\min_{x \in \mathbb{R}^n} f(x) + \psi(x)$$

if and only if for each $\alpha > 0$,

$$x^* = \operatorname{prox}_{(1/\alpha)\psi}(x^* - (1/\alpha)\nabla f(x^*)).$$

Hint: Use Exercises 2.24 and 2.25.

Exercise 2.27 (Subdifferentials are monotone)**.**
Let $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function, and let x, $y \in \operatorname{dom}(f)$. Show that

$$(g - h)^T (x - y) \geq 0 \quad \text{for all} \quad g \in \partial f(x), \quad h \in \partial f(y).$$

Note: This inequality generalizes (2.20).

Exercise 2.28 (Directional derivatives of proper convex functions)**.**
Let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex, and let $x \in \operatorname{rint}(\operatorname{dom}(f))$.
Establish the following statements.

1. For every $h \in \mathbb{R}^n$, the directional derivative $f'(x;h)$ of f at x in direction h, that is,
$$f'(x;h) := \lim_{t \to 0^+} \frac{f(x+th) - f(x)}{t},$$
exists and
$$f'(x;h) = \inf_{t>0} \frac{f(x+th) - f(x)}{t}.$$
Moreover the function $t \mapsto \frac{f(x+th)-f(x)}{t}$ is monotonically increasing on $(0,\infty)$.
2. The mapping $h \mapsto f'(x;h)$ is positively homogeneous, that is, for all $\lambda \geq 0$ and $h \in \mathbb{R}^n$, $f'(x;\lambda h) = \lambda f'(x;h)$.
3. The mapping $h \mapsto f'(x;h)$ is convex, and $f'(x;0) = 0$.
4. For all $h \in \mathbb{R}^n$, we have
$$-f'(x;-h) \leq f'(x;h).$$
5. For all $h \in \mathbb{R}^n$, we have
$$f(x+h) - f(x) \geq f'(x;h).$$
6. For each $h \in \mathbb{R}^n$,
$$f'(x;h) = \max_{g \in \partial f(x)} g^T h.$$

Exercise 2.29 (Convex matrix functions)**.**
The definition of a convex function as provided in Definition 2.1 canonically extends to sets X in spaces other than $\mathbb{R}^n$.
Establish the following statements.

1. The spectral norm $\|A\|_2$ of a matrix $A \in \mathbb{R}^{n\times m}$ is defined by $\|A\|_2 := \max\{\|Ax\|_2 \colon \|x\|_2 \leq 1\}$. Show that the spectral norm $\|\cdot\|_2$ on $\mathbb{R}^{n\times m}$ is convex.
2. Show that the maximum eigenvalue (mapping) $\lambda_{\max}(\cdot)$ on the space of symmetric $n \times n$ matrices is convex.

Exercise 2.30 (Empirical approximations of composite spectral risk measures [14, pp. 298–299])**.**
Let $F_1, \ldots, F_m$ be convex real-valued functions on a convex nonempty set $X \subset \mathbb{R}^n$. For each $x \in X$, let $F_{(1)}(x), \ldots, F_{(m)}(x)$ be the objective function values $F_1(x), \ldots, F_m(x)$ arranged in increasing order. In other words, $F_{(i)}(x)$ equals the ith-smallest value in the collection $F_1(x), \ldots, F_m(x)$. In statistics, $F_{(1)}(x), \ldots, F_{(m)}(x)$ is known as the order statistics of $F_1(x), \ldots, F_m(x)$.

Let $0 \leq q_1 \leq \cdots \leq q_m < \infty$ be constants (with $\sum_{i=1}^m q_i = 1$). Show that

$$g(x) \coloneqq \sum_{i=1}^m q_i F_{(i)}(x)$$

is convex.

Hints: (i) Establish the following version of the Hardy–Littlewood inequality: for all $q,\, p \in \mathbb{R}^m$,

$$\sum_{i=1}^m p_i q_i \leq \sum_{i=1}^m p_{(i)} q_{(i)},$$

where $q_{(1)} \leq \cdots \leq q_{(m)}$ is the order statistics of $q_1, \ldots, q_m$ and $p_{(1)} \leq \cdots \leq p_{(m)}$ is that of $p_1, \ldots, p_m$. (ii) Use convexity-preserving operations as discussed in (2.3).

Exercise 2.31 (Projection onto a polyhedral set)**.**
Let $z \in \mathbb{R}^n$ and let $a_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, where $n,\, m \in \mathbb{N}$ with $m < n$. We consider

$$\min_{x \in \mathbb{R}^n} (1/2)\|x - z\|_2^2 \quad \text{s.t.} \quad a_i^T x = 0, \quad i = 1, \ldots, m. \tag{2.22}$$

Let a_i^T be the ith row of the matrix $A \in \mathbb{R}^{m \times n}$. We assume that A has full rank.

1. Show that x^* solves (2.22) if and only if there exists $\mu^* \in \mathbb{R}^m$ such that

$$x^* - z + \sum_{i=1}^m \mu_i^* a_i = 0, \quad \text{and} \quad a_i^T x^* = 0, \quad i = 1, \ldots, m. \tag{2.23}$$

2. Show that (2.23) can be written as

$$x^* - z + A^T \mu^* = 0, \quad \text{and} \quad Ax^* = 0. \tag{2.24}$$

3. Show that $x^* \coloneqq z - A^T(AA^T)^{-1}Az$ solves (2.22).

   Hint: Use parts (a) and (b).

3 Convex Optimization: Duality, Optimality Conditions, and Saddle Points

We consider the mathematical optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X, \quad g_i(x) \le 0,\ i = 1,\dots,m, \quad h_j(x) = 0,\ j = 1,\dots,p, \tag{3.1}$$

where

- f is called *objective (function)*,
- $X \subset \mathbb{R}^n$ is called *domain*, and
- g_i for $i = 1,\dots,m$ are called *inequality constraints*, and h_j for $j = 1,\dots,p$ are called *equality constraints*.

If not stated otherwise, we assume that the objective function and the constraint functions are well-defined on X.

- A point $x \in \mathbb{R}^n$ is called *feasible* for (3.1) if $x \in X$, $g_i(x) \le 0$, $i = 1,\dots,m$, and $h_j(x) = 0$, $j = 1,\dots,p$. The *feasible set* of (3.1) is the set of all feasible points of (3.1).
- An inequality constraint $g_i(x)$ is called *active* at a feasible point x if $g_i(x) = 0$.
- A point $x^* \in \mathbb{R}^n$ is called *optimal solution* to (3.1) if it is feasible and $f(x^*) \le f(x)$ for each feasible point $x \in \mathbb{R}^n$ of (3.1).
- The *optimal value* f^* of (3.1) is defined by

$$f^* = \begin{cases} \inf_{x\in X, g(x)\le 0, h(x)=0} f(x) & \text{if (3.1) has feasible points} \\ +\infty & \text{if (3.1) has no feasible point}s \end{cases}$$

- The problem (3.1) is said to be *bounded from below* if the optimal value f^* is greater than $-\infty$, that is, if the objective function is bounded from below on the feasible set.
- The problem (3.1) is called *convex* if $X \subset \mathbb{R}^n$ is convex, $f, g_1, \dots, g_m$ are *real-valued* convex functions on X, and there are no equality constraints.

3.1 Convex Theorem on Alternative

Let $c \in \mathbb{R}$ be a scalar. We consider the nonlinear system

$$\begin{aligned} f(x) &< c, \\ g_i(x) &\le 0, \quad i = 1,\dots,m, \\ x &\in X. \end{aligned} \tag{3.2}$$

We are interested in certifying insolvability of the nonlinear system (3.2). If λ_i, $i = 1, \ldots, m$ are nonnegative weights such that the inequality

$$f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) < c$$

has no solutions in X, that is, if the system

$$\begin{aligned} \inf_{x \in X} \left[f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \right] &\geq c, \\ \lambda_i \geq 0, \quad i &= 1, \ldots, m, \end{aligned} \tag{3.3}$$

has a solution, then (3.2) has no solutions.

If the system (3.2) is convex and Slater's condition holds true, then we obtain a characterization of insolvability of (3.2). We say that the system (3.2) is *convex* if

1. X is a nonempty convex set, and
2. the functions f, $g_1, \ldots, g_m$ are real-valued convex functions on X.

We say that the *Slater condition* is satisfied for (3.2) if the subsystem

$$x \in X, \quad g_i(x) < 0, \quad i = 1, \ldots, m$$

has a solution. Moreover, we say that the *relaxed Slater condition* is satisfied for (3.2) if the functions $g_1, \ldots, g_k$ with $k \in \{0, 1, \ldots, m\}$ are affine linear, and there exists a point $x \in \operatorname{rint}(X)$ such that $g_i(\bar{x}) \leq 0$ for $i = 1, \ldots, k$ and $g_i(\bar{x}) < 0$ for $i = k+1, \ldots, m$. We allow for $k = 0$ which corresponds to the functions $g_1, \ldots, g_m$ being possibly nonaffine.

Let us talk a little bit about the geometric interpretation of the system (3.3). Let us define the vector-valued mapping $g \colon \mathbb{R}^n \to \mathbb{R}^m$ by

$$g(x) := \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}.$$

The system (3.3) is equivalent to

$$\begin{aligned} \begin{bmatrix} \lambda \\ 1 \end{bmatrix}^T \begin{bmatrix} g(x) \\ f(x) - c \end{bmatrix} &\geq 0, \quad \text{for all} \quad x \in X, \\ \lambda_i \geq 0, \quad i &= 1, \ldots, m. \end{aligned} \tag{3.4}$$

If there exists $\lambda \in \mathbb{R}^n$ satisfying the system (3.4), then the set

$$M := \{(g(x), f(x) - c) \colon x \in X\} \tag{3.5}$$

is contained in the halfspace $(\lambda, 1)^T z \geq 0$ with $z \in \mathbb{R}^m$ defined by a nonvertical hyperplane with normal vector $(\lambda, 1)$.

Theorem 3.1. *Let us consider the systems* (3.2) *and* (3.3) *with* $c \in \mathbb{R}$.

1. *If* (3.3) *has a solution, then* (3.2) *has no solution.*
2. *If* (3.2) *is convex, satisfies Slater's condition, and has no solution, then* (3.3) *has a solution.*

Proof. 1. We have established the first part before stating the theorem.
2. To establish the second part, we apply a separation theorem. We consider the sets

$$T = \{u \in \mathbb{R}^{m+1} \colon \exists x \in X \colon f(x) \leq u_0,\ g_i(x) \leq u_i,\ i = 1, \dots, m\}$$

and

$$S = \{u \in \mathbb{R}^{m+1} \colon u_0 < c,\ u_i \leq 0,\ i = 1, \dots, m\}.$$

The sets S and T are nonempty and convex and they do not intersect, as otherwise (3.2) would have a solution. (The potentially nonconvex set M defined in (3.5) is a subset of the convex set T.) Using the separation theorem, Theorem 1.34, we find that there exists a nonzero vector $a = (a_0, a_1, \dots, a_m) \in \mathbb{R}^{m+1}$ with

$$\sup_{u \in S} a^T u \leq \inf_{u \in T} a^T u.$$

Let us show that $a_i \geq 0$ for $i = 0, \dots, m$. If $a_i < 0$ for some $i \in \{0, \dots, m\}$, we would have $\sup_{u \in S} a^T u = +\infty$ which is impossible as T is nonempty. Since $a \geq 0$, we have

$$\sup_{u \in S} a^T u = \sup_{\substack{u_0 < c, \\ u_i \leq 0, i = 1, \dots, m}} [a_0 u_0 + \dots + a_m u_m] = a_0 c.$$

Hence

$$\inf_{u \in T} a^T u \geq a_0 c.$$

For each $u \in T$, we have

$$a_0 u_0 + a_1 u_1 + \dots + a_m u_m \geq a_0 c$$

and there exists a point $x \in X$ with $f(x) \leq u_0$ and $g_i(x) \leq u_i$ for $i = 1, \dots, m$. Hence

$$\inf_{x \in X} [a_0 f(x) + a_1 g_1(x) + \dots + a_m g_m(x)] \geq a_0 c. \tag{3.6}$$

Let us now show that $a_0 > 0$. Suppose that $a_0 = 0$. Then $(a_1, \dots, a_m) \neq 0$ and

$$\inf_{x \in X} [a_1 g_1(x) + \dots + a_m g_m(x)] \geq 0.$$

Since $a_i > 0$ for at least one $i \in \{1, \dots, m\}$ and Slater's condition ensures the existence of a point $\bar{x} \in X$ with $g_i(\bar{x}) < 0$ for $i = 1, \dots, m$, this inequality cannot be true.

Since $a_0 > 0$, we can divide (3.6) by a_0. Defining $\lambda_i = a_i / a_0$ for $i = 1, \dots, m$, we obtain $c \leq \inf_{x \in X} f(x) + \sum_{i=1}^m \lambda_i g_i(x)$.

□

We now generalize Theorem 3.1 to allow for the relaxed Slater's condition rather than Slater's condition.

Theorem 3.2. *Let us consider the systems* (3.2) *and* (3.3) *with* $c \in \mathbb{R}$.

1. *If* (3.3) *has a solution, then* (3.2) *has no solution.*
2. *If* (3.2) *is convex, satisfies the relaxed Slater condition, and has no solution, then* (3.3) *has a solution.*

Proof. See Exercise 3.6. □

3.2 Lagrange Duality

We consider the optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X, \quad g_i(x) \leq 0, \quad i = 1, \ldots, m. \tag{3.7}$$

We associate with (3.7) the *Lagrange function* $L \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined by

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \tag{3.8}$$

and its *Lagrange dual problem*

$$\sup_{\lambda \geq 0} \underline{L}(\lambda), \quad \text{where} \quad \underline{L}(\lambda) \coloneqq \inf_{x \in X} L(x, \lambda). \tag{3.9}$$

The function $\underline{L}$ is called *dual function.* Let us denote the optimal value of (3.7) by Opt(P) and that of its dual problem (3.9) by Opt(D). We refer to (3.7) also as *primal problem.*

Theorem 3.3 (Weak and strong duality)**.**

1. (weak duality) *For every* $\lambda \geq 0$, *we have* $\underline{L}(\lambda) \leq \mathrm{Opt(P)}$. *In particular,*
$$\mathrm{Opt(D)} \leq \mathrm{Opt(P)}.$$
2. (strong duality) *If the primal problem* (3.7) *is convex, bounded from below, and satisfies Slater's condition, then the dual problem* (3.9) *has a solution, and*
$$\mathrm{Opt(D)} = \mathrm{Opt(P)}.$$

Proof. 1. If the primal problem (3.7) is infeasible, then we have $\mathrm{Opt(P)} = +\infty$. If x is feasible for the primal problem (3.7) and $\lambda \geq 0$, then we have $L(x, \lambda) \leq f(x)$. Hence for $\lambda \geq 0$,

$$\begin{aligned}
\underline{L}(\lambda) &= \inf_{x \in X} L(x, \lambda) \\
&\leq \inf_{x \text{ is feasible for the primal problem}} L(x, \lambda) \\
&\leq \inf_{x \text{ is feasible for the primal problem}} f(x) \\
&= \mathrm{Opt(P)}
\end{aligned}$$

Taking suprema with respect to $\lambda \geq 0$, we obtain $\mathrm{Opt(D)} \leq \mathrm{Opt(P)}$.

2. We apply the convex theorem on alternative, Theorem 3.1, to establish the existence of a solution to the dual problem (3.9) and $\mathrm{Opt(D)} \geq \mathrm{Opt(P)}$. The feasible set of the primal problem (3.7) is nonempty, as Slater's condition holds. Combined with the fact that the primal problem is bounded from below, we find that $\mathrm{Opt(P)}$ is finite. The system

$$f(x) < \mathrm{Opt(P)}, \quad g_i(x) \leq 0,\ i = 1, \ldots, m, \quad x \in X$$

has no solutions. Theorem 3.1 ensures the existence of nonnegative weights $\lambda_i^* \geq 0$ for $i = 1, \ldots, m$ such that

$$\inf_{x \in X} L(x, \lambda^*) \geq \mathrm{Opt(P)}.$$

Hence

$$\underline{L}(\lambda^*) \geq \mathrm{Opt(P)}.$$

Combined with weak duality, we obtain

$$\mathrm{Opt(D)} = \underline{L}(\lambda^*) = \mathrm{Opt(P)}.$$

Hence $\lambda^* \geq 0$ solves the dual problem (3.9).

□

If the dual optimal value $\mathrm{Opt(D)}$ is bounded from below, that is, $\mathrm{Opt(D)} > -\infty$, then the difference

$$\mathrm{Opt(P)} - \mathrm{Opt(D)}$$

is called *duality gap*. Theorem 3.3 ensures that the duality gap is always nonnegative. Many algorithms for convex programming us termination criteria based on the duality gap.

We compute the Lagrangian dual for a linear program.

Example 3.4. We consider the linear program (LP)

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax \leq b,$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$. The Lagrangian function L is given by

$$L(x, \lambda) = c^T x + \lambda^T (Ax - b) = -b^T \lambda + x^T (A^T \lambda + c).$$

Hence, the (Lagrangian) dual function becomes

$$\underline{L}(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda) = -b^T \lambda + \inf_{x \in \mathbb{R}^n} x^T (A^T \lambda + c) = \begin{cases} -b^T \lambda & \text{if} \quad (A^T \lambda + c) = 0, \\ -\infty & \text{if} \quad (A^T \lambda + c) \neq 0. \end{cases}$$

Therefore, we can write the Lagrangian dual problem as

$$\max_{\lambda \in \mathbb{R}^m} -b^T \lambda \quad \text{s.t.} \quad A^T \lambda = -c, \quad \lambda \geq 0.$$

For convex programming problems, Theorem 3.3 provides conditions sufficient for strong duality.

Whether strong duality holds true or not can depend on the representation of the constraints.

Example 3.5. We consider the nonconvex optimization problem

$$\min_{x\in\mathbb{R}} -x^2 \quad \text{s.t.} \quad |x| - 1 \leq 0,$$

with $f(x) \coloneqq -x^2$ and $g(x) \coloneqq |x| - 1$. The optimal value of the primal problem is -1. For the Lagrange dual function, we obtain

$$\underline{L}(\lambda) = \inf_{x\in\mathbb{R}} -x^2 + \lambda(|x| - 1) = -\infty.$$

Hence $\mathrm{Opt(D)} = -\infty$. However, we can show that strong holds true if we consider instead the equivalent primal problem

$$\min_{x\in\mathbb{R}} -x^2 \quad \text{s.t.} \quad x^2 - 1 \leq 0,$$

with $f(x) \coloneqq -x^2$ and $g(x) \coloneqq x^2 - 1$.

Let us note that the primal problem (3.7) can be written in terms of the Lagrange function. We have

$$\mathrm{Opt(P)} = \inf_{x\in X} \sup_{\lambda\geq 0} L(x, \lambda). \tag{3.10}$$

Indeed, we have for all $x \in X$,

$$\sup_{\lambda\geq 0} L(x,\lambda) = \begin{cases} f(x) & \text{if} \quad g_i(x) \leq 0,\ i = 1, \dots, m, \\ +\infty & \text{otherwise.} \end{cases}$$

3.3 Saddle Points of the Lagrange Function

We turn our attention to the minimax problem defined in (3.10) and discuss the saddle point form of optimality conditions for the optimization problem (3.7). A point $(x^*, \lambda^*) \in X \times \mathbb{R}^m_+$ is called a *saddle point* of the Lagrange function L defined in (3.8) over $X \times \mathbb{R}^m_+$ if

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*) \quad \text{for all} \quad (x, \lambda) \in X \times \mathbb{R}^m_+.$$

Let us related saddle points of the Lagrange function to global solutions of the (3.7) and its Lagrange dual (3.9).

Theorem 3.6.

1. (Characterization of saddle points and sufficient optimality conditions) *Let* $(x^*, \lambda^*) \in X \times \mathbb{R}^m_+$ *be a point. Then the following statements are equivalent.*

a) $(x^*, \lambda^*) \in X \times \mathbb{R}^m_+$ *is a saddle point of the Lagrange* L *over* $X \times \mathbb{R}^m_+$.

b) x^* *is a global solution to* (3.7), λ^* *is a global solution to* (3.9), *and* $\text{Opt(D)} = \text{Opt(P)}$.

2. (Necessary optimality conditions) *If* $x^* \in X$ *is optimal for* (3.7), (3.7) *is* convex, *and satisfies Slater's condition, then there exists* $\lambda^* \geq 0$ *such that* (x^*, λ^*) *is a saddle point of the Lagrange function* L *over* $X \times \mathbb{R}^m_+$.

Proof. 1. "$\Rightarrow$" Since (x^*, λ^*) is a saddle point

$$L(x^*, \lambda^*) = \inf_{x \in X} L(x, \lambda^*) \leq \sup_{\lambda \geq 0} \inf_{x \in X} L(x, \lambda) \leq \inf_{x \in X} \sup_{\lambda \geq 0} L(x, \lambda) \leq \sup_{\lambda \geq 0} L(x^*, \lambda) = L(x^*, \lambda^*)$$

Hence $\text{Opt(D)} = \text{Opt(P)}$. We have $\sup_{\lambda \geq 0} L(x^*, \lambda) = +\infty$ if x^* is infeasible and $\sup_{\lambda \geq 0} L(x^*, \lambda) = f(x^*)$ otherwise. Since (x^*, λ^*) is a saddle point, we have $\sup_{\lambda \geq 0} L(x^*, \lambda) \leq L(x^*, \lambda^*) < \infty$. Hence x^* is feasible. Moreover

$$\underline{L}(\lambda^*) = f(x^*) = L(x^*, \lambda^*).$$

Now weak duality (see Theorem 3.3) implies that x^* and λ^* are optimal.

"$\Leftarrow$" We have $\text{Opt(D)} = \underline{L}(\lambda^*) = \inf_{x \in X} L(x, \lambda^*)$, and $\text{Opt(P)} = f(x^*) = \sup_{\lambda \geq 0} L(x^*, \lambda)$. Hence

$$L(x^*, \lambda^*) \leq f(x^*) = \text{Opt(P)} = \text{Opt(D)} = \underline{L}(\lambda^*) = \sup_{\lambda \geq 0} L(x^*, \lambda) = \inf_{x \in X} L(x, \lambda^*) \leq L(x^*, \lambda^*).$$

The first inequality results from $\lambda^* \geq 0$ and $g_i(x^*) \leq 0$, $i = 1, \ldots, m$. The last inequality follows from $x^* \in X$. We deduce

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*) \quad \text{for all} \quad (x, \lambda) \in X \times \mathbb{R}^m_+.$$

2. Theorem 3.3 ensures the existence of $\lambda^* \geq 0$ with

$$f(x^*) = \underline{L}(\lambda^*) = \inf_{x \in X} \left\{ f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) \right\}. \tag{3.11}$$

Combined with the feasibility of x^*, we obtain

$$f(x^*) = \inf_{x \in X} \left\{ f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) \right\} \leq f(x^*) + \sum_{i=1}^m \lambda_i^* g_i(x^*) \leq f(x^*).$$

Consequently, $\lambda_i^* g_i(x^*) = 0$, $i = 1, \ldots, m$. Combined with $g_i(x^*) \leq 0$, we obtain

$$f(x^*) = L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^m \underbrace{\lambda_i^* g_i(x^*)}_{=0} \geq f(x^*) + \sum_{i=1}^m \lambda_i \underbrace{g_i(x^*)}_{\leq 0} = L(x^*, \lambda) \quad \text{for all} \quad \lambda \geq 0.$$

Moreover, (3.11) ensures $f(x^*) \leq L(x, \lambda^*)$ for all $x \in X$. Combining these two inequalities, we obtain the assertion.

$\square$

3.4 Karush–Kuhn–Tucker Optimality Conditions

We discuss first-order necessary and sufficient optimality conditions for (3.7). We recall that the *normal cone* $N_X(x)$ of X at $x \in X$ is the set

$$N_X(x) = \{\, h \in \mathbb{R}^n \colon\ h^T(x-y) \geq 0 \text{ for all } y \in X \,\}.$$

Theorem 3.7. *Let* (3.7) *be a convex optimization problem, let* $x^* \in \mathbb{R}^n$, *and let* f, $g_1, \ldots, g_m$ *be differentiable at* x^*.

1. (Sufficiency) *If there exist Lagrange multipliers* $\lambda_i^* \in \mathbb{R}$, $i = 1, \ldots, m$, *such that*

$$\begin{gathered} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) \in N_X(x^*), \\ x^* \in X, \quad \lambda_i^* \geq 0, \quad g_i(x^*) \leq 0, \quad \lambda_i^* g_i(x^*) = 0, \quad i = 1, \ldots, m, \end{gathered} \tag{3.12}$$

   then x^* *is a solution to* (3.7).
2. (Necessity and sufficiency) *If furthermore the Slater's condition holds for* (3.7), *then the KKT conditions in* (3.12) *are necessary and sufficient for* x^* *to be an optimal solution to* (3.7).

The conditions in (3.12) are referred to as *KKT conditions* for (3.7). The conditions $\lambda_i^* g_i(x^*) = 0$, $i = 1, \ldots, m$, are referred to as *complementary conditions*.

Proof of Theorem 3.7. 1. In light of Theorem 3.6 it suffices to show that (x^*, λ^*) is a saddle point of the Lagrange function L over $X \times \mathbb{R}^m_+$. We have $\nabla_x L(x^*, \lambda^*) \in N_X(x^*)$. Combined with the convexity of $L(\cdot, \lambda^*)$ and of X, and Theorem 2.18 with (2.8), we obtain $L(x, \lambda^*) \geq L(x^*, \lambda^*)$ for all $x \in X$. The complementary conditions and the feasbility of x^* ensure $L(x^*, \lambda^*) \geq L(x^*, \lambda)$ for all $\lambda \geq 0$. Putting together the pieces, we find that (x^*, λ^*) is a saddle point of L over $X \times \mathbb{R}^m_+$. Now Theorem 3.6 ensures that x^* is optimal.

2. Theorem 3.6 ensures that there exists a saddle point (x^*, λ^*) of L over $X \times \mathbb{R}^m_+$. Hence the complementary conditions hold and $L(x, \lambda^*) \geq L(x^*, \lambda^*)$ for all $x \in X$. Hence x^* is a minimizer of $L(\cdot, \lambda^*)$ over X. Using Theorem 2.18 with (2.8), we obtain $\nabla_x L(x^*, \lambda^*) \in N_X(x^*)$.

□

3.5 Saddle points*

It is possible to consider a more general form of the minimax problem than that in (3.10): Let $X \subset \mathbb{R}^n$ and $\Lambda \subset \mathbb{R}^m$ be nonempty sets and $F \colon X \times \Lambda \to \mathbb{R}$ be a real-valued function on $X \times \Lambda$. This setting gives rise to two optimization problems:

$$\begin{aligned} \mathrm{Opt}(\mathrm{P}') &= \inf_{x \in X} \sup_{\lambda \in \Lambda} F(x, \lambda), \\ \mathrm{Opt}(\mathrm{D}') &= \sup_{\lambda \in \Lambda} \inf_{x \in X} F(x, \lambda). \end{aligned}$$

A short verification shows that

$$\sup_{\lambda\in\Lambda}\inf_{x\in X} F(x,\lambda) \leq \inf_{x\in X}\sup_{\lambda\in\Lambda} F(x,\lambda). \tag{3.13}$$

Indeed, for all $x \in X$ and $\lambda \in \Lambda$, we have

$$F(x,\lambda) \leq \sup_{\lambda\in\Lambda} F(x,\lambda).$$

Minimizing over $x \in X$, we obtain

$$\inf_{x\in X} F(x,\lambda) \leq \inf_{x\in X}\sup_{\lambda\in\Lambda} F(x,\lambda).$$

Now maximization over $\lambda \in \Lambda$ ensures (3.13). The inequality (3.13) is sometimes called *max-min inequality*. If

$$\sup_{\lambda\in\Lambda}\inf_{x\in X} F(x,\lambda) = \inf_{x\in X}\sup_{\lambda\in\Lambda} F(x,\lambda),$$

then we say that f (and X and Λ) satisfy the *saddle-point* property. We refer to the tuple $(x^*,\lambda^*) \in X \times \Lambda$ as *saddle point* of F over $X \times \Lambda$ if

$$F(x^*,\lambda) \leq F(x^*,\lambda^*) \leq F(x,\lambda^*) \quad \text{for all} \quad (x,\lambda) \in X \times \Lambda.$$

We observe that the notion of a saddle point of a Lagrangian function L over $X \times \mathbb{R}^m_+$ is a special case of the above definition. Theorem 3.6 can be extended to this more general setting.

3.6 Exercises

Exercise 3.1.
Compute the minimizer of the linear function

$$f(x) = c^T x$$

over the set

$$V_p = \left\{ x \in \mathbb{R}^n \colon \sum_{i=1}^n |x_i|^p \leq 1 \right\},$$

where $1 < p < \infty$.

Hint: Use the KKT conditions established in Theorem 3.7.

Exercise 3.2.
Let $a_1, \ldots, a_n > 0$ and let $\alpha, \beta > 0$. Solve the optimization problem

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^n \frac{a_i}{x_i^\alpha} \quad \text{s.t.} \quad x > 0, \quad \sum_{i=1}^n x_i^\beta \leq 1.$$

Hint: Use the variable transformation $y_i = x_i^\beta$ to obtain a convex optimization problem, and use the KKT conditions to solve the optimization problem.

Exercise 3.3.
Consider the optimization problem

$$\max_{x \in \mathbb{R}^n, t \in \mathbb{R}^n} \xi^T x + \tau t + \ln(t^2 - x^T x) \quad \text{s.t.} \quad (x,t) \in X \coloneqq \{(x,t) \colon t > \sqrt{x^T x}\},$$

where $\xi \in \mathbb{R}^n$ and $\tau \in \mathbb{R}$ are parameters. Is the problem convex? For which choices of parameters is the problem solvable? What is the optimal value? Is the optimal value convex in the parameters?

Exercise 3.4.
Consider the optimization problem

$$\max_{x,y \in \mathbb{R}} ax + by + \ln(\ln(y) - x) + \ln(y) \quad \text{s.t.} \quad (x,y) \in X \coloneqq \{(x,y) \in \mathbb{R}^2 \colon y > \exp(x)\},$$

where a, $b \in \mathbb{R}$ are parameters. Is the problem convex? For which choices of parameters is the problem solvable? What is the optimal value? Is the optimal value convex in the parameters?

Exercise 3.5 (Inconsistency of strict convex inequalities)**.**
Let $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$ be convex functions and let X be a nonempty convex set.

1. Consider the optimization problem

$$\min_{(x,t)\in\mathbb{R}^n\times\mathbb{R}} t \quad \text{s.t.} \quad x \in X,\ g_i(x) \le t,\ i = 1, \dots, m. \tag{3.14}$$

Show that the corresponding dual function $\underline{L}$ is given by

$$\underline{L}(\lambda) = \begin{cases} \inf_{x\in X} \sum_{i=1}^m \lambda_i g_i(x) & \text{if} \quad \sum_{i=1}^m \lambda_i = 1, \\ -\infty & \text{if} \quad \sum_{i=1}^m \lambda_i \neq 1, \end{cases}$$

2. Show that the system

$$g_i(x) < 0, \quad i = 1, \dots, m$$

has no solution in X if and only if there exists $\lambda_i \in \mathbb{R}$, $i = 1, \dots, m$ with

$$\lambda_i \ge 0, \quad i = 1, \dots, m, \quad \sum_{i=1}^m \lambda_i = 1,$$

$$\sum_{i=1}^m \lambda_i g_i(x) \ge 0, \quad \text{for all} \quad x \in X.$$

Exercise 3.6.
Prove Theorem 3.2.

Exercise 3.7 (Strong duality of the trust-region subproblem*)**.**
We consider the trust-region subproblem

$$\min_{x\in\mathbb{R}^n} x^T Ax + 2b^T x \quad \text{s.t.} \quad \|x\|_2^2 \le 1, \tag{3.15}$$

where $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. This problem arises in trust-region methods as a subproblem. We define $f(x) := x^T Ax + 2b^T x$ and $g(x) = \|x\|_2^2 - 1$. Let $L(x, \lambda) = f(x) + \lambda g(x)$ be its Lagrangian.

Show that strong duality holds for (3.15). You can provide your own proof or solve the following subproblems. Let $\lambda_{\min} := \lambda_{\min}(A)$ be the minimum eigenvalue of A.

1. Let $\lambda_{\min}(A)$ be nonnegative. Show that strong duality holds.

For the remainder, let $\lambda_{\min}(A) < 0$.

2. Establish $\min_{\|x\|_2^2\leq 1}\{\, x^\top Ax+2b^\top x \,\} = \min_{\|x\|_2^2\leq 1}\{\, x^\top(A-\lambda_{\min}I)x+2b^\top x+\lambda_{\min}1 \,\}$.
   Hints: You can use without proof the following facts.
   (i) Since $\lambda_{\min} < 0$, each solution x^* to(3.15) satisfies $\|x^*\|_2 = 1$.
   (ii) Since the minimum eigenvalue of $A-\lambda_{\min}I$ is zero, $\min_{\|x\|_2^2\leq 1}\{\, x^\top(A-\lambda_{\min}I)x+2b^\top x \,\}$ has a solution y^* with $\|y^*\|_2 = 1$.
3. Prove that $\min_{\|x\|_2^2\leq 1}\{\, x^\top(A-\lambda_{\min}I)x+2b^\top x+\lambda_{\min} \,\} \leq \sup_{\lambda\in\mathbb{R}_{\geq 0}} \inf_{x\in\mathbb{R}^n} L(x,\lambda)$, where L is the Lagrangian function corresponding to (3.15).
4. Combine the above computations with weak duality to deduce the strong duality for(3.15).

Exercise 3.8 (Projection onto ℓ_1-unit ball)**.**
Let $x \in \mathbb{R}^n$. Show that the solution $P(x)$ to

$$\min_{v\in\mathbb{R}^n} (1/2)\|v-x\|_2^2 \quad \text{s.t.} \quad \|v\|_1 \leq 1$$

is given by $P_C(x) = x$ if $\|x\|_1 \leq 1$ and by

$$P(x)_k = \operatorname{sign}(x_k)\max\{|x_k|-\lambda^*, 0\} = \begin{cases} x_k-\lambda^* & \text{if} \quad x_k > \lambda^* \\ 0 & \text{if} \quad -\lambda^* \leq x_k \leq \lambda^* \\ x_k+\lambda^* & \text{if} \quad x_k < -\lambda^* \end{cases}$$

otherwise, where λ^* is the solution to

$$\sum_{k=1}^{n} \max\{|x_k|-\lambda^*, 0\} = 1.$$

Hint: Use Theorem 3.6.

Exercise 3.9 (Projection onto probability simplex)**.**
Let $x \in \mathbb{R}^n$. Show that the solution $P(x)$ to

$$\min_{v\in\mathbb{R}^n} (1/2)\|v-x\|_2^2 \quad \text{s.t.} \quad \|v\|_1 = 1, \quad v \geq 0$$

is given by

$$P(x)_k = \max\{x_k-\lambda, 0\},$$

where λ^* is the solution to

$$\sum_{k=1}^{n} \max\{x_k-\lambda, 0\} = 1.$$

Hints: (i) The projection problem is equivalent to

$$\min_{v\in\mathbb{R}^n} (1/2)\|v-x\|_2^2 + I_{\mathbb{R}^n_{\geq 0}}(v) \quad \text{s.t.} \quad \|v\|_1 = 1,$$

where $I_{\mathbb{R}^n_{\geq 0}}(v) = +\infty$ if $v \geq 0$ and $I_{\mathbb{R}^n_{\geq 0}}(v) = 0$ otherwise. (ii) Use Theorem 3.6.

Exercise 3.10 (proximal mapping)**.**
A function $f\colon \mathbb{R}^n \to (-\infty,\infty]$ is lower semicontinuous if for each $r \in \mathbb{R}$, the level set $\{x \in \mathbb{R}^n \colon f(x) \leq r\}$ is closed.

For a proper, convex, lower semicontinuous function f, the proximal mapping is defined by

$$\operatorname{prox}_f(x) \coloneqq \operatorname{argmin}_{v\in\mathbb{R}^n} f(v) + (1/2)\|x-v\|_2^2.$$

This mapping is well-defined (the optimization problem has a unique solution). If f is the indicator function of a nonempty, closed, convex set S, then $\operatorname{prox}_f(x)$ becomes the projection of x onto S.

Show the following implications. Let $t \geq 0$.

1. If $f(x) = \|x\|_2$, then

$$\operatorname{prox}_{tf}(x) = \max\{1 - t/\|x\|_2, 0\}x = \begin{cases} (1-t/\|x\|_2)x & \text{if} \quad \|x\|_2 \geq t \\ 0 & \text{otherwise} \end{cases}$$

2. If $A \in \mathbb{R}^{n\times n}$ is symmetric and positive definite, and $f(x) = (1/2)x^TAx + b^Tx + c$, then

$$\operatorname{prox}_{tf}(x) = (I+tA)^{-1}(x-tb).$$

Exercise 3.11 (Trust-region composite subproblems*)**.**
We consider

$$\min_{x\in\mathbb{R}^n} x^TAx + 2b^Tx + \gamma\|x\|_1 \quad \text{s.t.} \quad \|x\|_2^2 \leq 1, \tag{3.16}$$

where $\gamma \geq 0$, $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix, and $b \in \mathbb{R}^n$ is a vector. We are interested in whether strong duality holds for (3.16). Let $L(x,\lambda) \coloneqq x^TAx+2b^Tx+\gamma\|x\|_1+\lambda(\|x\|_2^2-1)$ be the Lagrangian of (3.16).

- Show that if $n=1$, $\gamma=1$, $A=-2$, and $b=0$, then the optimal value of (3.16) is -1 and dual optimal value is -2.
- Show that if $n=1$, $\gamma=1$, $A=-1$, and $b=2$, then the optimal value of (3.16) is -2 and dual optimal value is -2.

- Let λ^* be a solution to the Lagrangian dual to (3.16), and let $x(\lambda^*)$ be a solution to

$$\min_{x \in \mathbb{R}^n} L(x, \lambda^*). \tag{3.17}$$

Does $x(\lambda^*)$ solve (3.16)?

- Show that the optimal value of

$$\min_{x \in \mathbb{R}^n, X \in \mathbb{S}^n} A \bullet X + 2b^T x + \|x\|_1 \quad \text{s.t.} \quad A \bullet I \leq 1, \quad X \succcurlyeq xx^T$$

is less than or equal to that of (3.16).

4 Optimality Conditions for Nonlinear Optimization Problems

We consider the nonlinear optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_i(x) \le 0,\ i = 1, \ldots, m, \quad h_j(x) = 0,\ j = 1, \ldots, p. \tag{4.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function, $g_i : \mathbb{R}^n \to \mathbb{R}$ are the inequality constraints, and $h_j : \mathbb{R}^n \to \mathbb{R}$ are the equality constraints. We define the mappings $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ by

$$g(x) \coloneqq \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}, \quad \text{and} \quad h(x) \coloneqq \begin{bmatrix} h_1(x) \\ \vdots \\ h_p(x) \end{bmatrix}.$$

This notation is useful when establishing theoretical statements, but less so for analytic computations.

A point x is called feasible for (4.1) if $g(x) \le 0$ and $h(x) = 0$. For a feasible point x, we define the *set of active inequality constraints* $\mathcal{A}(x)$ by

$$\mathcal{A}(x) \coloneqq \{\, i \in \{1, \ldots, m\} \colon g_i(x) = 0 \,\}.$$

The equality constraints $h_j(x) = 0$, $j = 1, \ldots, p$, are active by definition.

We introduce further terminology. A point $x^* \in \mathbb{R}^n$ is a *(global) solution* to (4.1) if it is feasible for (4.1) and $f(x^*) \le f(x)$ for all feasible x. A point $x^* \in \mathbb{R}^n$ is a *local solution* to (4.1) if it is feasible for (4.1) and if there exists $r > 0$ such that $f(x^*) \le f(x)$ for all for all feasible x with $\|x - x^*\|_2 \le r$. A point $x^* \in \mathbb{R}^n$ is a *strict local solution* to (4.1) if it is feasible for (4.1) and if there exists $r > 0$ such that $f(x^*) < f(x)$ for all feasible $x \ne x^*$ with $\|x - x^*\|_2 \le r$. Figure 4.1 provides an illustration.

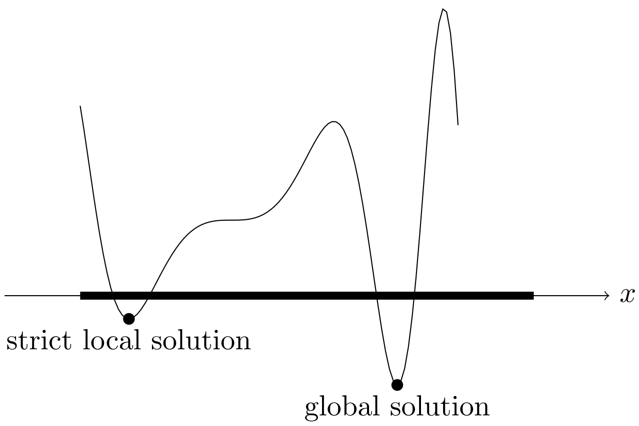

Figure 4.1: The graph of a function, the feasible set (thick black line), and the strict local and global solutions/minimizers.

Given a local solution x^* to (4.1), we derive first-order and second-order necessary conditions that are satisfied at x^*. Moreover, we derive second-order sufficient optimality conditions.

4.1 First-order optimality conditions

In the following theorem, we state first-order optimality conditions for the constrained problem (4.1). They are referred to as first-order conditions, as they involve first-order derivatives. The validity of these first-order optimality conditions are based on a condition on the constraints in (4.1).

Definition 4.1. *A feasible point x for* (4.1) *is called* regular *if the gradients $\nabla g_i(x)$, $i \in \mathcal{A}(x)$, and $\nabla h_i(x)$, $i = 1, \dots, p$, are linearly independent.*

In other words, a feasible point (4.1) is regular if the gradients of the active constraints are linearly independent. A necessary condition for a point to be regular is that the number of active constraints is less than or equal to n. Indeed, if the number of active constraints exceeds n, then the gradients of active constraints cannot be linearly independent.

The *Lagrange function* $L\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated with (4.1) is defined by

$$L(x, \lambda, \mu) \coloneqq f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x).$$

Theorem 4.2. *Let f, g, and h be continuously differentiable and let x^* be a local solution to* (4.1)*. Suppose that x^* is regular. Then there exist Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and*

$\mu^* \in \mathbb{R}^p$ *such that*

$$\begin{aligned} \nabla_x L(x^*, \lambda^*, \mu^*) &= 0, \\ h_j(x^*) &= 0,\ j = 1, \dots, p, \\ \lambda_i^* \geq 0, \quad g_i(x^*) \leq 0, \quad g_i(x^*)\lambda_i^* &= 0,\ i = 1, \dots, m. \end{aligned} \tag{4.2}$$

Theorem 4.2 remains valid if no inequality constraints are present in (4.1). In this case, we omit all terms involving λ^* and g in (4.2). Moreover, Theorem 4.2 remains valid if no equality constraints are present in (4.1). In this case, we omit all terms involving μ^* and h in (4.2).

Note that the conditions $g_i(x^*) \leq 0$, $i = 1, \dots, m$, and $h_j(x^*) = 0$, $j = 1, \dots, p$, are a direct consequence of x^* being a local solution to (4.1), as a local solution must be feasible. We have $\lambda_i^* = 0$ if $g_i(x^*) < 0$. In other words, if the ith inequality constraint is inactive, then the corresponding Lagrange multiplier must be zero.

If f, g, and h are differentiable, then we have

$$\nabla_x L(x, \lambda, \mu) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \mu_i \nabla h_i(x).$$

Using this gradient formula, we find that the conditions in (4.2) imply that $-\nabla f(x^*)$ is a linear combination of the gradients $\nabla g_i(x^*)$, $i \in \mathcal{A}(x^*)$ and $\nabla h_j(x^*)$, $j = 1, \dots, p$. If there are no equality constraints in (4.1), then the conditions in (4.2) imply that that $-\nabla f(x^*)$ is a conic combination of the gradients $\nabla g_i(x^*)$, $i \in \mathcal{A}(x^*)$. Figure 4.2 provides an illustration.

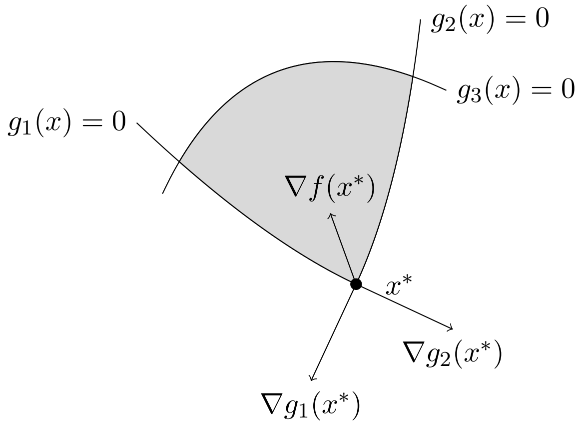

Figure 4.2: The gray region is the feasible set. The point x^* is feasible and $-\nabla f(x^*)$ is a conic combination of $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$. Hence x^* is a KKT point. The point x^* is regular, as $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$ are not linearly dependent.

We introduce standard terminology.

Definition 4.3. 1. *The conditions in* (4.2) *are referred to as* Karush–Kuhn–Tucker (KKT) conditions *of* (4.1).

2. *A feasible point* $x^* \in \mathbb{R}^n$ *is called* KKT point *of* (4.1) *if there exist* $\lambda^* \in \mathbb{R}^m$ *and* $\mu^* \in \mathbb{R}^p$ *such that the conditions in* (4.2) *hold.*
3. *If* (x^*, λ^*, μ^*) *satisfy the KKT conditions, then* (x^*, λ^*, μ^*) *is referred to as* KKT triple.
4. *The condition*

$$g_i(x^*)\lambda_i^* = 0, \ i = 1, \ldots, m.$$

is referred to as complementarity condition *or* complementary condition*.*

Using this terminology, Theorem 4.2 ensures that local solutions to (4.1) are KKT points, provided that they satisfy a constraint qualification. However, the KKT conditions are *not* sufficient optimality conditions for nonconvex optimization problems. If a point x^* is a KKT point of (4.1) and (4.1) is nonconvex, it may not be a local solution. We illustrate this on an example.

Example 4.4. We consider

$$\min_{x \in \mathbb{R}} -x^2 \quad \text{s.t.} \quad x^2 - 1 \le 0.$$

In order to formulate this problem as an instance of (4.1), we define $f(x) = -x^2$ and $g(x) = x^2 - 1$. The global minimizers to the problem are ± 1 and $x^* = 0$ is the global maximizer. We have for all $\lambda^* \in \mathbb{R}$,

$$f'(x^*) + \lambda^* g'(x^*) = -2x^* + 2\lambda^* x^* = 0.$$

Moreover, $g(x^*) = -1$ and hence $(x^*, 0)$ is a KKT point of (4.1). However, it is not even a local minimizer.

Before we discuss a proof of Theorem 4.2, we consider an optimization problem with unique solution which fails to be a KKT point.

Example 4.5. We consider the one-dimensional optimization problem

$$\min_{x \in \mathbb{R}} x \quad \text{s.t.} \quad x^2 = 0. \tag{4.3}$$

Since $x^* = 0$ is the only feasible point of (4.3), it is the unique solution to (4.3). In order to formulate (4.3) as an instance of (4.1), we define $f(x) \coloneqq x$ and $h(x) \coloneqq x^2$. We have $f'(x) = 1$ and $h'(x) = 2x$. Hence for all $\mu \in \mathbb{R}$,

$$f'(x^*) + \mu h'(x^*) = 1.$$

Therefore, $x^* = 0$ is not a KKT point of (4.3). Since $h'(x^*) = 0$, x^* is not a regular point of (4.3).

4.1.1 Proof of the first-order necessary conditions

Before we establish the first-order necessary conditions in Theorem 4.2, we derive first-order optimality conditions for unconstrained minimization problems.

Proposition 4.6. *Let $U \subset \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be differentiable. If $x^* \in U$ is a local minimizer of f, then $\nabla f(x^*) = 0$.*

To establish Proposition 4.6, we can reduce ourselves to a "one-dimensional" setting. If x^* is a local minimizer of f, then for each direction $d \in \mathbb{R}^n$, $t^* = 0$ is a local minimizer of $\phi(t) := f(x^* + td)$. Hence $0 = \phi'(0) = \nabla f(x^*)^T d$. Choosing $d = \nabla f(x^*)$ yields $\nabla f(x^*) = 0$.

We present a somewhat alternative proof of Proposition 4.6. It provides a motivation for some technical steps in the proof of Theorem 4.2.

Proof of Proposition 4.6. Let $d \in \mathbb{R}^n$ and let $(t_k) \subset (0, \infty)$ be a sequence converging to 0 as $k \to \infty$. Since x^* is a local minimizer of f, there exists $r > 0$ such that $f(x^*) \le f(x)$ for all $x \in \mathbb{R}^n$ with $\|x - x^*\|_2 \le r$. Let us define $x^k := x^* + t_k d$. For all sufficiently large $k \in \mathbb{N}$, we have $\|x^k - x^*\|_2 = t_k \|d\|_2 \le r$ and

$$f(x^k) \ge f(x^*).$$

Hence

$$\frac{f(x^* + t_k d) - f(x^*)}{t_k} \ge 0.$$

Taking limits as $k \to \infty$, we find that

$$\nabla f(x^*)^T d \ge 0.$$

This inequality is valid for all $d \in \mathbb{R}^n$. Let us choose $d = -\nabla f(x^*)$. Then we have

$$-\nabla f(x^*)^T \nabla f(x^*) \ge 0.$$

In other words, $\|\nabla f(x^*)\|_2^2 \le 0$. Hence $\nabla f(x^*) = 0$. $\square$

We prepare our proof of Theorem 4.2. To derive first-order necessary optimality condition for a local solution x^*, we would like to use ideas from the proof of Proposition 4.6. In the proof of Proposition 4.6, we constructed for arbitrary $d \in \mathbb{R}^n$, a perturbation, $x^k = x^* + t_k d$, of x^* that approaches x^* as k increases. Therefore, we have $f(x^k) \ge f(x^*)$ for sufficiently large k. The special perturbation, $x^k = x^* + t_k d$, has allowed us to make use of the differentiability of f to obtain the inequality $\nabla f(x^*)^T d \ge 0$. We would like to apply a similar approach for establishing Theorem 4.2. However, this approach raises at least one question: given $d \in \mathbb{R}^n$, why should $x^* + td$ be feasible for (4.1) for any $t > 0$? In general, the points $x^* + td$ may be infeasible for (4.1). Therefore, we need a more technical constructions of perturbations x^k of x^*. This construction exploits the regularity of x^*. Furthermore, if $x^* + td$ would be feasible for (4.1) for all $t > 0$ and all

$d \in \mathbb{R}^n$, then we would have $\nabla f(x^*) = 0$, as the proof of Proposition 4.6 shows. We will need to restrict ourselves to directions $d \in \mathbb{R}^n$ from some smaller set than $\mathbb{R}^n$.

For a feasible point x for (4.1), we define

$$T_\ell(g,h,x) \coloneqq \{d \in \mathbb{R}^n \colon \nabla g_i(x)^T d \le 0,\ i \in \mathcal{A}(x), \quad \nabla h_j(x)^T d = 0,\ j = 1,\dots,p\}. \tag{4.4}$$

This set is a nonempty closed cone and is sometimes referred to as *the set of linearized feasible directions*. If the problem formulation (4.1) does not involve inequality constraints, we write $T_\ell(h,x)$ and if it does not involve equality constraints, we write $T_\ell(g,x)$. We have come across this cone already in Example 2.20. If the inequality constraints are given by $g_i(x) = a_i^T x - b_i$ and we have no equality constraints in (4.1), then $T_\ell(g,x)$ equals the radial cone $T_X(x)$ of the set $X = \{x \in \mathbb{R}^n \colon a_i^T x \le b_i,\ i = 1,\dots,m\}$; see Example 2.20.

Figure 4.3 provides an illustration.

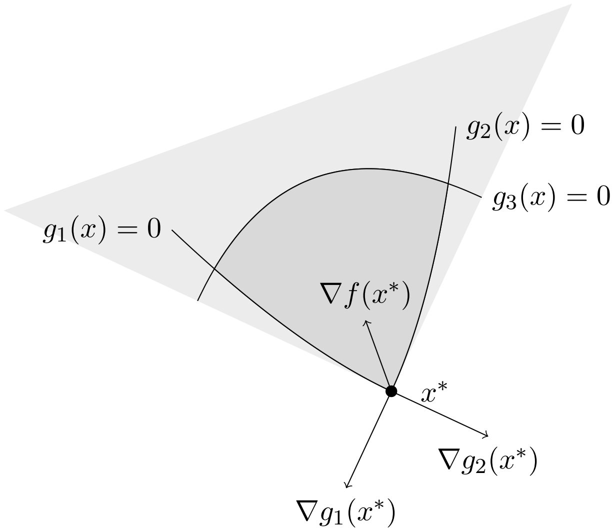

Figure 4.3: The gray region is the feasible set. The light gray area is the shifted cone $x^* + T_\ell(g,x^*)$. For directions $d \in T_\ell(g,x^*)$ with small norm, the points $x^* + d$ are “close” to being feasible.

Its name can be motivated by considering first-order expansions of g and h: for $d \in T_\ell(g,h,x^*)$ and small $t > 0$, we have

$$h_i(x^* + td) \approx h_i(x^*) + t\nabla h_i(x^*)^T d = t\nabla h_i(x^*)^T d = 0, \quad i = 1,\dots,p.$$

If $i \in \mathcal{A}(x^*)$, then

$$g_i(x^* + td) \approx g_i(x^*) + t\nabla g_i(x^*)^T d = t\nabla g_i(x^*)^T d \le 0.$$

If $i \notin \mathcal{A}(x^*)$, we have $g_i(x^*) < 0$. Since g_i is differentiable and hence continuous, we have $g_i(x^* + td) < 0$ for small $t > 0$.

Given $d \in T_\ell(g,h,x^*)$, we construct a sequence (x^k) of feasible points for (4.1) such that (x^k) converges to x^* and $d^k := k^{-1}(x^k - x^*)$ converges to d. Therefore, we eventually have $f(x^k) \geq f(x^*)$. It will allow us to show that $\nabla f(x^*)^T d \geq 0$ for all $d \in T_\ell(g,h,x^*)$. Figure 4.4 depicts an illustration of this construction.

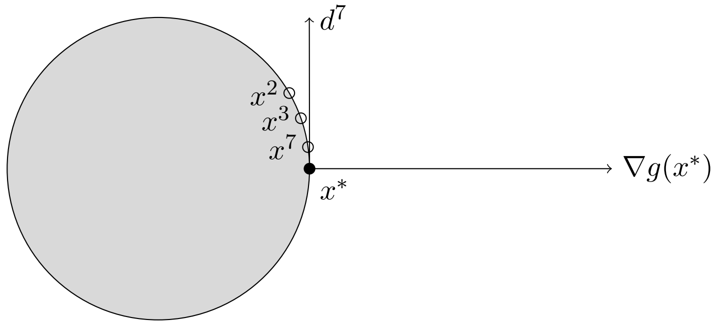

Figure 4.4: The gray region is the feasible set defined by the inequality constraint $g(x) = x_1^2 + x_2^2 - 1$. For $x^* = (1,0)$, we have $\nabla g(x^*) = (2,0) \neq 0$. So x^* is a regular point. The vector $d = (0,1)$ is contained in $T_\ell(g,h,x^*)$. The sequence $x^k = \begin{bmatrix} \sqrt{1-(1/k)^2} \\ 1/k \end{bmatrix}$ is feasible, as $g(x^k) = 0$, and approaches $\bar{x}$. Moreover $d^k := k(x^k - \bar{x})$ approaches d as $k \to \infty$.

Lemma 4.7. *If x^* is feasible for (4.1) and regular, then for each $d \in T_\ell(g,h,x^*)$, there exist feasible points x^k of (4.1) with $x^k \to x^*$ as $k \to \infty$ such that $k(x^k - x^*) \to d$ as $k \to \infty$.*

The proof of Lemma 4.7 is based on an application of the implicit function theorem stated next. We use $F_x(x,y)$ to denote the partial derivative of F with respect to x.

Theorem 4.8 (Implicit function theorem)**.** *Let $F : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^n$ be q-times continuously differentiable with $q \in \mathbb{N}$ and let $(\bar{x},\bar{y}) \in \mathbb{R}^n \times \mathbb{R}^s$ be a point with $F(\bar{x},\bar{y}) = 0$. Suppose that the matrix $F_x(\bar{x},\bar{y}) \in \mathbb{R}^{n\times n}$ is invertible. Then there exist open sets $U \subset \mathbb{R}^n \times \mathbb{R}^s$ and $W \subset \mathbb{R}^s$ with $(\bar{x},\bar{y}) \in U$ and $\bar{y} \in W$ having the following properties:*

1. *For every $y \in W$, there exists a unique $x \in \mathbb{R}^n$ such that*
$$(x,y) \in U \quad \text{and} \quad F(x,y) = 0. \tag{4.5}$$
2. *If this x is defined to be $G(y)$, then the mapping $G : W \to \mathbb{R}^n$ is q-times continuously differentiable with $G(\bar{y}) = \bar{x}$ and*
$$G'(\bar{y}) = -F_x(\bar{x},\bar{y})^{-1} F_y(\bar{x},\bar{y}).$$

Using the implicit function theorem, we are ready to establish Lemma 4.7. We denote by $|\mathcal{A}(x^*)|$ the number of active inequality constraints and by $g_{\mathcal{A}(x^*)}$ the component function g_i of g with $i \in \mathcal{A}(x^*)$. In other words, $g_{\mathcal{A}(x^*)}$ consists of those component functions of g corresponding to active inequality constraints. We recall that the derivative H' of a mapping H is the transpose of its gradient ∇H. In other words, $H'(y) = \nabla H(y)^T$.

Proof of Lemma 4.7. We use the implicit function theorem to establish the assertion. Let $A \in \mathbb{R}^{n \times |\mathcal{A}(x^*)|+p}$ be the matrix with columns $\nabla g_i(x^*)$, $i \in \mathcal{A}(x^*)$, and $\nabla h_i(x^*)$, $i = 1, \dots, p$. By assumption A has full rank. Let $B \in \mathbb{R}^{n \times n - |\mathcal{A}(x^*)|-p}$ be a matrix such that its columns form a basis of the null space of A^T. Since x^* is regular, the square matrix

$$\begin{bmatrix} A^T \\ B^T \end{bmatrix}$$

is invertible.

Let $d \in T_\ell(g, h, x^*)$. We define $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ by

$$F(x,t) = \begin{bmatrix} g_{\mathcal{A}(x^*)}(x) - t\nabla g_{\mathcal{A}(x^*)}(x^*)^T d \\ h(x) - t\nabla h(x^*)^T d \\ B^T(x - x^* - td) \end{bmatrix}$$

To apply the implicit function theorem, Theorem 4.8, we verify its hypotheses.

1. F is continuously differentiable,
2. $F(x^*, 0) = 0$, and
3. we have

$$F_x(x,0) = \begin{bmatrix} g'_{\mathcal{A}(x^*)}(x) \\ h'(x) \\ B^T \end{bmatrix}.$$

Hence

$$F_x(x^*,0) = \begin{bmatrix} A^T \\ B^T \end{bmatrix}$$

Since the hypotheses of the implicit function theorem are satisfied, there exist open sets $U \subset \mathbb{R}^n \times \mathbb{R}$ and $W \subset \mathbb{R}$ such that $(\bar{x}, 0) \in U$ and $0 \in W$ with the following properties:

1. For every $t \in W$, there exists a unique $x(t) \in \mathbb{R}^n$ such that $F(x(t), t) = 0$.
2. $x(t)$ is continuously differentiable on W, $x(0) = x^*$, $x'(0) = -\nabla_x F(x^*, 0)^{-1} \nabla_t F(\bar{x}, 0)$.

Since $x(t)$ is continuously differentiable on W, it is continuous. Hence we have $x(t) \to x^*$ as $t \to 0$. Moreover, we have

$$\frac{x(t) - x^*}{t} \to x'(0).$$

Let us compute $x'(0)$. We have

$$F_t(\bar{x},0) = \begin{bmatrix} -\nabla g_{\mathcal{A}(x^*)}(x^*)^T d \\ -\nabla h(x^*)^T d \\ -B^T d \end{bmatrix} = \begin{bmatrix} -A^T d \\ -B^T d \end{bmatrix} = -\begin{bmatrix} A^T \\ B^T \end{bmatrix} d.$$

Hence

$$x'(0) = -F_x(x^*,0)^{-1} F_t(\bar{x},0) = \begin{bmatrix} A^T \\ B^T \end{bmatrix}^{-1} \begin{bmatrix} A^T \\ B^T \end{bmatrix} d = d.$$

Since $F(x(t),t)=0$ and $d \in T_\ell(g,h,x^*)$, we obtain for $t \in W$,

$$\begin{aligned} g_{\mathcal{A}(x^*)}(x(t)) &= t\nabla g_{\mathcal{A}(x^*)}(x^*)^T d \leq 0, \\ h(x(t)) &= t\nabla h(x^*)^T d = 0. \end{aligned} \tag{4.6}$$

If $i \notin \mathcal{A}(x^*)$, then we have $g_i(x^*) < 0$. Combined with the continuity of g_i and $x(t) \to x^*$ as $t \to 0$, we find that $g_i(x(t)) < 0$ for all sufficiently small $t \in W$. Therefore $x(t)$ is feasible for (4.1), provided that $t \in W$ is sufficiently small.

Now, we can construct the points x^k. Since $W \subset \mathbb{R}$ is open and $0 \in W$, the set W is an open interval about 0. So we have $1/k \in W$ for all sufficiently large $k \in \mathbb{N}$. The above computations show that $x(1/k)$ is feasible for (4.1) for all sufficiently large $k \in \mathbb{N}$. For all such k's, we define $x^k = x(1/k)$. Otherwise, we choose $x^k = x^*$. Our computations show that the x^k's are feasible and that $k(x^k - x^*) \to d$ as $k \to \infty$. □

Using Lemma 4.7, we can establish a first-order necessary optimality condition. This condition is key to establishing the validity of the KKT conditions.

Lemma 4.9. *Under the hypotheses of Theorem* 4.2*, we have* $\nabla f(x^*)^T d \geq 0$ *for all* $d \in T_\ell(g,h,x^*)$.

The assertion of Lemma 4.9 is interesting. It implies that the vector $d^* = 0$ is a solution to the linear program (LP)

$$\min_{d\in\mathbb{R}^n} \nabla f(x^*)^T d \quad \text{s.t.} \quad \nabla g_i(x^*)^T d \leq 0,\ i \in \mathcal{A}(x^*), \quad \nabla h_j(x^*)^T d = 0,\ j = 1,\dots,p. \tag{4.7}$$

Proof of Lemma 4.9. Let $d \in T_\ell(g,h,x^*)$. Using Lemma 4.7, there exist sequences (x^k) with $x^k \to x^*$ as $k \to \infty$ and x^k is feasible for (4.1) and $(t_k) \subset (0,\infty)$ with $t_k \to 0$ as $k \to \infty$ such that $d^k \coloneqq (x^k - x^*)/t_k$ converges to d as $k \to \infty$.

Since x^* is a local solution, there exists $r > 0$ such that $f(x) \geq f(x^*)$ for all feasible points x with $\|x - x^*\|_2 \leq r$. As $x^k \to x^*$ as $k \to \infty$, there exists a natural number K such that $\|x^k - x^*\| \leq r$ for all $k \geq K$.

Using a first-order Taylor's expansion, we obtain

$$f(x^k) - f(x^*) = \nabla f(x^*)^T (x^k - x^*) + o(\|x^k - x^*\|_2),$$

where $o(t)$ is a function with $o(t)/t \to 0$ as $t \to 0$. For $k \geq K$, we have $f(x^k) \geq f(x^*)$. Hence for $k \geq K$,

$$0 \leq \nabla f(x^*)^T (x^k - x^*) + o(\|x^k - x^*\|_2).$$

Dividing by $t_k > 0$ and using $d^k = (x^k - x^*)/t_k$, we obtain

$$0 \leq \nabla f(x^*)^T d^k + \frac{o(\|x^k - x^*\|_2)}{\|x^k - x^*\|_2} \cdot \|d^k\|_2.$$

Taking limits as $k \to \infty$, we obtain

$$\nabla f(x^*)^T d \geq 0.$$

□

We are now ready to establish Theorem 4.2. Our proof applies the homogeneous Farkas lemma to the assertion of Lemma 4.9. An alternative approach is to apply LP duality to the LP (4.7).

Proof of Theorem 4.2. We apply the homogeneous Farkas lemma to the assertion of Lemma 4.9. The assertion of Lemma 4.9 is: $\nabla f(x^*)^T d \geq 0$ for all $d \in T_\ell(g,h,x^*)$. Using the definition of $T_\ell(g,h,x^*)$, we obtain $\nabla f(x^*)^T d \geq 0$ for all $d \in \mathbb{R}^n$ with $\nabla g_i(x^*)^T d \leq 0$ for $i \in \mathcal{A}(x^*)$ and $\nabla h_i(x^*)^T d \leq 0$ and $-\nabla h_i(x^*)^T d \leq 0$ for $i = 1,\dots,p$. The homogeneous Farkas lemma ensures the existence of nonnegative vectors $\lambda \in \mathbb{R}^{|\mathcal{A}(x^*)|}$ and $\mu_+ \in \mathbb{R}^p$ and $\mu_- \in \mathbb{R}^p$ such that

$$\nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} (\mu_+ - \mu_-)_i \nabla h_i(x^*) = 0.$$

We define $\lambda_i^* = \lambda_i$ if $i \in \mathcal{A}(x^*)$ and $\lambda_i^* = 0$ otherwise, and $\mu^* = \mu_+ - \mu_-$. These definitions ensure the complementary condition $\lambda_i^* g_i(x^*) = 0$, $i = 1,\dots,m$. To summarize, (x^*, λ^*, μ^*) satisfy the KKT conditions. □

4.1.2 Further interpretations

Our proof of Theorem 4.2 relies on the technical statements, Lemmas 4.7 and 4.9. As discussed after the formulation of Lemma 4.9, the hypotheses of Theorem 4.2 ensures that the direction $d^* = 0$ solves the LP (4.7). It turns out that under the hypotheses of Theorem 4.2, the KKT point x^* considered in Theorem 4.2 is a solution to the LP

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & f(x^*) + \nabla f(x^*)^T (x - x^*) \\ \text{s.t.} \ & g_i(x^*) + \nabla g_i(x^*)^T (x - x^*) \leq 0, \ i = 1,\dots,m, \\ & h_j(x^*) + \nabla h_j(x^*)^T (x - x^*) = 0, \ j = 1,\dots,p. \end{aligned} \tag{4.8}$$

Since this optimization problem is an LP, we can also use optimality conditions for LPs to establish the KKT conditions.

Let us now provide a relationship between the LPs (4.7) and (4.8).

Lemma 4.10. *Let $x^* \in \mathbb{R}^n$ feasible for* (4.1). *The point $d^* = 0$ is a solution to the LP* (4.7) *if and only if x^* is a solution to the LP* (4.8).

Proof. "⇒" Let $d^* = 0$ be a solution to the LP (4.7) and let $x \in \mathbb{R}^n$ be feasible for (4.8). We have to show that

$$f(x^*) + \nabla f(x^*)^T (x - x^*) \geq f(x^*).$$

Let us define $d = x - x^*$. The point d is feasible for the LP (4.7). Hence

$$\nabla f(x^*)^T d \geq \nabla f(x^*)^T d^* = 0.$$

Hence x^* is a solution to (4.8).

"$\Leftarrow$" Now let x^* be a solution to (4.8) and let $d \in \mathbb{R}^n$ be feasible for (4.7). We have to show that $d^* = 0$ solves (4.7). Let us consider the points $x_t = x^* + td$ and show that x_t is feasible for (4.8), provided that $t > 0$ is sufficiently small. We have for all $t > 0$,

$$0 = h_j(x^*) + \nabla h_j(x^*)^T(x_t - x^*) = 0 + t\nabla h_j(x^*)^T d,\ j = 1, \ldots, p.$$

If $i \in \mathcal{A}(x^*)$, then we have $g_i(x^*) = 0$ and

$$0 \geq g_i(x^*) + \nabla g_i(x^*)^T(td) = \nabla g_i(x^*)^T(td).$$

If $i \notin \mathcal{A}(x^*)$, then we have $g_i(x^*) < 0$ and hence

$$g_i(x^*) + \nabla g_i(x^*)^T(td) \leq 0$$

for all sufficiently small $t > 0$. Therefore x_t is feasible for (4.8) for some $t > 0$. We have

$$f(x^*) + t\nabla f(x^*)^T d = f(x^*) + \nabla f(x^*)^T(x_t - x^*) \geq f(x^*).$$

Hence $d^* = 0$ solves the LP (4.7). □

4.1.3 Constraint Qualifications

In the first-order necessary optimality conditions provided in Theorem 4.2, we may replace the sentence

"Suppose that x^* is a regular point."

by

"Suppose that x^* satisfies a constraint qualification."

We define the technical notion "constraint qualification." Let $M \subset \mathbb{R}^n$ be a nonempty set and let $\bar{y} \in M$. The tangent cone of M at $\bar{y}$ is defined by

$$T_M(\bar{y}) := \{d \in \mathbb{R}^n \colon \exists \eta_k > 0,\ y^k \in X,\ y^k \to \bar{y},\ \eta_k(y^k - \bar{y}) \to d\}.$$

The tangent cone of the feasible set of (4.8) is the linearized tangent cone defined in (4.4) (this fact is very helpful for graphical illustrations). For a cone $K \subset \mathbb{R}^n$, we define its polar cone K° by

$$K^\circ := \{v \in \mathbb{R}^n \colon v^T d \leq 0 \quad \text{for all} \quad d \in K\}.$$

Let X be the feasible set of (4.1). The Guignard constraint qualification (GCQ) is satisfied at $\bar{x} \in X$ if

$$T_\ell(g, h, \bar{x})^\circ = T(X, \bar{x})^\circ.$$

Any condition implying GCQ is called a *constraint qualification.* We provide a non-exhaustive list of *constraint qualifications* (CQs).

- The constraints g and h are affine linear
- The constraints g are convex and h is affine linear, and there exists a point $y \in \mathbb{R}^n$ with $g(y) < 0$ and $h(y) = 0$ (Slater's condition).
- The function g is concave and h is affine.
- Mangasarian–Fromovitz CQ (MFCQ): A feasible point x satisfies the MFCQ if
  1. $\nabla h(x)$ as full column rank, and
  2. there exists $d \in \mathbb{R}^n$ with
  $$\nabla g_i(x)^T d < 0, \quad i \in \mathcal{A}(x), \quad \nabla h(x)^T d = 0.$$
- Generalized MFCQ: A feasible point x satisfies the generalized MFCQ if
  1. $\nabla h(x)$ as full column rank, or h is affine linear, and
  2. there exists $d \in \mathbb{R}^n$ with
  $$\nabla g_i(x)^T d < 0, \quad i \in \mathcal{A}(x), \quad \nabla h(x)^T d = 0.$$

All of these conditions ensure the assertion of Lemma 4.9, that is, if x^* is a local solution to (4.1) and x^* satisfies one of the above constraint qualifications, then $\nabla f(x^*)^T d \geq 0$ for all directions $d \in T_\ell(g, h, x^*)$.

Roughly speaking, constraint qualifications ensure that the feasible set of the nonlinear problem (4.1) provides a "good" approximation of feasible set of the linearized optimization problem (4.8) for all points x in a neighborhood of x^*. Figure 4.5 depicts an example where the feasible set of (4.1) provides a poor approximation to that of (4.7).

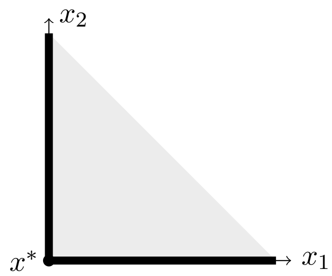

Figure 4.5: The thick black lines illustrate the feasible of an instance of the nonlinear problem (4.1) with $g_1(x) = -x_1$, $g_2(x) = -x_2$, $g_3(x) = x_1 x_2$, and the feasible point $x^* = 0$. The light gray area is the cone $T_\ell(g, x^*)$.

*If g_i, $i = 1, \ldots, m$, are convex and h_j, $j = 1, \ldots, p$, are affine linear, and there exists a feasible point, then the generalized MFCQ and Slater's condition are equivalent. To establish this fact, let x be feasible. Suppose that there exists $y \in \mathbb{R}^n$ with $g(y) < 0$ and $h(y) = 0$. We show that $d = y - x$ satisfies the conditions of the generalized MFCQ. Since g_i are convex, the gradient inequality implies for $i \in \mathcal{A}(x)$,

$$0 > g_i(y) \geq g_i(x) + \nabla g_i(x)^T (y - x) = \nabla g_i(x)^T d.$$

Since h_j are affine linear, we have $0 = h(y) = h(x) + \nabla h(x)^T(y - x) = \nabla h(x)^T d$. Consequently, $d = y - x$ satisfies the conditions in the generalized MFCQ.

Now, let x be a feasible point that satisfies the generalized MFCQ. Then there exists $d \in \mathbb{R}^n$ with $\nabla g_i(x)^T d < 0$, $i \in \mathcal{A}(x)$ and $\nabla h(x)^T d = 0$. For any $t > 0$, we have $h(x + td) = h(x) + t\nabla h(x)^T d = h(x) = 0$, as h is affine and x is feasible. Let $i \in \mathcal{A}(x)$ and define $\phi(t) = g_i(x + td)$. We have $\phi(0) = g_i(x) = 0$ and $\phi'(0) = \nabla g_i(x)^T d < 0$. Hence for all sufficiently small $t > 0$, we have $\phi(t) < 0$ and therefore $g_i(x + td) < 0$. If $i \notin \mathcal{A}(x)$, then $g_i(x) < 0$. Since g_i is continuous, we have $g_i(x+td) < 0$ for all sufficiently small $t > 0$. We conclude that $x + td$ satisfies the Slater's condition for some $t > 0$.

4.2 Second-order necessary optimality conditions

Theorem 4.2 provides first-order necessary optimality conditions. In this section, we establish second-order necessary optimality conditions.

The formulation of these second-order necessary conditions require us to introduce a cone related to that in (4.4). Let $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ be a vector. To formulate the second-order conditions, we define the *critical cone*

$$T_+(g, h, x, \lambda) = T_\ell(g, h, x) \cap \{\, d \in \mathbb{R}^n \colon \nabla g_i(x)^T d = 0 \ \text{ if } \ i \in \mathcal{A}(x) \ \text{ and } \ \lambda_i > 0 \,\}. \tag{4.9}$$

The set $T_+(g, h, x, \lambda)$ is a closed nonempty cone. Using the definition of $T_\ell(g, h, x)$ provided in (4.4), we can write

$$T_+(g, h, x, \lambda) = \left\{ d \colon \nabla g_i(x)^T d \begin{cases} = 0 & \text{if } i \in \mathcal{A}(x) \text{ and } \lambda_i > 0, \\ \leq 0 & \text{if } i \in \mathcal{A}(x) \text{ and } \lambda_i = 0, \end{cases} \ \nabla h_j(x)^T d = 0, \ j = 1, \ldots, p \right\}.$$

Now we state second-order necessary optimality conditions.

Theorem 4.11. *Let f, g, and h be twice continuously differentiable, and let x^* be a local solution to* (4.1)*. Suppose that x^* is regular. Then there exists Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that (x^*, λ^*, μ^*) is a KKT triple of* (4.1) *and*

$$d^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) d \geq 0 \quad \textit{for all} \quad d \in T_+(g, h, x^*, \lambda^*).$$

The second-order necessary optimality conditions may be used to show that a KKT point of (4.1) is *not* a local solution of (4.1). See Exercise 4.2.

The proof of Theorem 4.11 consists of two steps: Fix $d \in T_+(g, h, x^*, \lambda^*)$.

1. Construct feasible points x^k for (4.1) with $(x^k - x^*)/t_k \to d$ as $k \to \infty$, where $t_k > 0$ with $t_k \to 0$ as $k \to \infty$. This uses regularity of x^*.
2. Demonstrate $d^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) d \geq 0$. This uses, among other things, feasibility of x^k, local optimality of x^*, and the fact that (x^*, λ^*, μ^*) is a KKT triple of (4.1).

Before establishing second-order necessary conditions for (4.1), we derive second-order necessary conditions for unconstrained minimization problems in Proposition 4.12. This serves two purposes. First, the proof of Theorem 4.11 uses ideas that of Proposition 4.12. Second, we utilize Proposition 4.12 in our proof of Theorem 4.11.

Proposition 4.12. *Let $U \subset \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be twice continuously differentiable. If $x^* \in U$ is a local minimizer of f, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite.*

Proof. Proposition 4.6 ensures $\nabla f(x^*) = 0$. Let $d \in \mathbb{R}^n$ be a vector. We define $\phi(t) := f(x^* + td)$. The function ϕ is twice continuously differentiable, $t^* = 0$ is a local minimizer of ϕ, and $\phi'(0) = 0$. Using a second-order Taylor's expansion, we have as $t \to 0$,

$$\phi(t) = \phi(0) + \phi'(0)t + (1/2)\phi''(0)t^2 + o(t^2).$$

Using $\phi'(0) = 0$ and $\phi(t) \geq \phi(0)$ for all sufficiently small $t > 0$, we obtain

$$0 \leq (1/2)\phi''(0)t^2 + o(t^2).$$

Dividing by t^2 and taking limits as $t \to 0$, we obtain $\phi''(0) \geq 0$. It remains to note that $\phi''(0) = d^T \nabla^2 f(x^*) d$. □

Proof of Proposition 4.12. The existence of Lagrange multipliers and the fact that (x^*, λ^*, μ^*) is a KKT triple of (4.1) is implied by Theorem 4.2.

Let $d \in T_+(g, h, x^*, \lambda^*)$. Let us define the set $\mathcal{A}_0 = \{i \colon \nabla g_i(x^*)^T d = 0\}$. Since $d \in T_+(g, h, x^*, \lambda^*)$, we can use the technique in the proof of Lemma 4.7 to construct feasible points x^k with $(x^k - x^*)/t_k \to d$ as $k \to \infty$, where $t_k > 0$ with $t_k \to 0$ as $k \to \infty$. Because of our construction of the points x^k, we obtain from (4.6),

$$h_j(x^k) = t_k \nabla h_j(x^*)^T d, \ j = 1, \dots, p, \quad \text{and} \quad g_i(x^k) = t_k \nabla g_j(x^*)^T d, \ i \in \mathcal{A}_0.$$

If $i \notin \mathcal{A}_0$, then we either have $g_i(x^*) < 0$ or $i \in \mathcal{A}(x^*)$ with $\nabla g_i(x^*)^T d < 0$. If $g_i(x^*) < 0$, then the complementary condition yields $\lambda_i^* = 0$. If $i \in \mathcal{A}(x^*)$ with $\nabla g_i(x^*)^T d < 0$, then $d \in T_+(g, h, x^*, \lambda^*)$ yields $\lambda_i^* = 0$. Combined with $d \in T_+(g, h, x^*, \lambda^*)$, we have

$$\begin{aligned}
L(x^k, \lambda^*, \mu^*) &= f(x^k) + \sum_{i=1}^m \lambda_i^* g_i(x^k) + \sum_{j=1}^p \mu_j^* h_j(x^k) \\
&= f(x^k) + \sum_{i \in \mathcal{A}_0} \lambda_i^* g_i(x^k) + \sum_{j=1}^p \mu_j^* h_j(x^k) \\
&= f(x^k) - t_k \sum_{i \in \mathcal{A}_0} \lambda_i^* \nabla g_j(x^*)^T d - t_k \sum_{j=1}^p \mu_j^* \nabla h_j(x^*)^T d \\
&= f(x^k).
\end{aligned}$$

Using a second-order Taylor's expansion of $L(\cdot, \lambda^*, \mu^*)$ about x^*, we obtain

$$\begin{aligned}
L(x^k, \lambda^*, \mu^*) = {} & L(x^*, \lambda^*, \mu^*) + \nabla_x L(x^*, \lambda^*, \mu^*)^T (x^k - x^*) \\
& + (1/2)(x^k - x^*)^T \nabla_{xx} L(x^*, \lambda^*, \mu^*)(x^k - x^*) + o(\|x^k - x^*\|_2^2).
\end{aligned}$$

The complementary condition further implies $f(x^*) = L(x^*, \lambda^*, \mu^*)$ and we have $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$. Putting together the pieces, we find that

$$f(x^k) = f(x^*) + 1/2)(x^k - x^*)^T \nabla_{xx} L(x^*, \lambda^*, \mu^*)(x^k - x^*) + o(\|x^k - x^*\|_2^2).$$

Since x^k is feasible and x^* is a local solution, we have $f(x^k) \geq f(x^*)$ for all sufficiently large k. We obtain for those k's,

$$0 \leq (1/2)(x^k - x^*)^T \nabla_{xx} L(x^*, \lambda^*, \mu^*)(x^k - x^*) + o(\|x^k - x^*\|_2^2).$$

Dividing by t_k^2 and defining $d^k := (x^k - x^*)/t_k$, we obtain

$$0 \leq (1/2)(d^k)^T \nabla_{xx} L(x^*, \lambda^*, \mu^*) d^k + \frac{o(\|x^k - x^*\|^2)}{\|x^k - x^*\|^2} \|d^k\|_2^2.$$

Using $d^k \to d$ as $k \to \infty$ and taking limits as $k \to \infty$, we obtain

$$0 \leq (1/2) d^T \nabla_{xx} L(x^*, \lambda^*, \mu^*) d.$$

□

4.3 Second-order sufficient optimality conditions

We provide second-order conditions, which assert local optimality of KKT points.

Theorem 4.13. *Let f, g, and h be twice continuously differentiable, and let (x^*, λ^*, μ^*) be a KKT triple of* (4.1). *If*

$$d^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) d > 0 \quad \textit{for all} \quad d \in T_+(g, h, x^*, \lambda^*) \setminus \{0\},$$

then x^* is a strict local solution to (4.1).

The second-order sufficient optimality conditions provided in Theorem 4.13 may be used to show that a KKT point is a local solution. However, they do not assert *global* optimality.

Our approach is as before. We first establish second-order sufficient optimality conditions for unconstrained problems.

Proposition 4.14. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. If $x^* \in \mathbb{R}^n$, $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite, then x^* is a strict local minimizer of f.*

Proposition 4.14 does not assert *global* optimality. Moreover if the second-order sufficient optimality conditions (see Proposition 4.14) are satisfied at x^*, then x^* may not be a global minimizer.

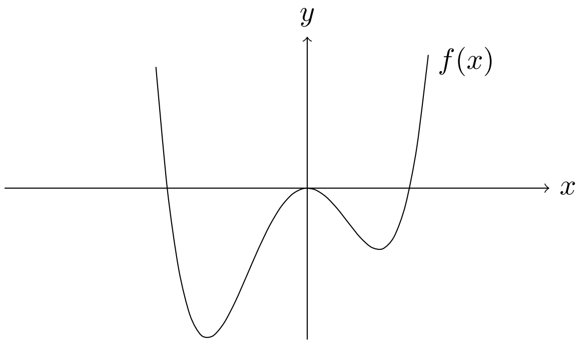

Figure 4.6: The graph of $f(x) := x^4 + x^3 - 10x^2$. The function f has two local minmizers and both satisfy the second-order sufficient optimality conditions. The local minmizer on the positive part of the horizontal axis is not a global minimzer.

Proof of Proposition 4.14. Suppose that x^* is not a strict local minimizer of f. Then there exists a sequence (x^k) with $x^k \neq x^*$, $x^k \to x^*$ as $k \to \infty$, and $f(x^k) \leq f(x^*)$. We define

$$d^k := \frac{x^k - x^*}{\|x^k - x^*\|_2}.$$

We have $\|d^k\|_2 = 1$. Therefore, the sequence (d^k) has an accumulation point d with $\|d\|_2 = 1$. We may assume that $d^k \to d$ as $k \to \infty$. Using a second-order Taylor's expansion, we obtain

$$f(x^k) = f(x^*) + \nabla f(x^*)^T(x^k - x^*) + (1/2)(x^k - x^*)\nabla^2 f(x^*)(x^k - x^*) + o(\|x^k - x^*\|_2^2).$$

Using $f(x^k) \leq f(x^*)$ and $\nabla f(x^*)$, we obtain

$$0 \geq (1/2)(x^k - x^*)\nabla^2 f(x^*)(x^k - x^*) + o(\|x^k - x^*\|_2^2).$$

Dividing by $\|x^k - x^*\|_2^2$, using the definition of d^k and $d^k \to d$ as $k \to \infty$, we obtain $d^T\nabla^2 f(x^*)d \leq 0$. Since $d \neq 0$, this contradicts our assumption that $\nabla^2 f(x^*)$ is positive definite. □

Proof of Theorem 4.13. We use ideas from the proof of Proposition 4.14. Suppose that x^* is not a strict local solution to (4.1). Then there exists a sequence (x^k) with x^k feasible for (4.1), $x^k \neq x^*$, $x^k \to x^*$ as $k \to \infty$, and $f(x^k) \leq f(x^*)$. We define

$$d^k := \frac{x^k - x^*}{\|x^k - x^*\|_2}.$$

We have $\|d^k\|_2 = 1$. Therefore, the sequence (d^k) has an accumulation point d with $\|d\|_2 = 1$. We may assume that $d^k \to d$ as $k \to \infty$.

Our first goal is to show that $d \in T_+(g, h, x^*, \lambda^*)$. We have

$$0 \geq \frac{f(x^k) - f(x^*)}{\|x^k - x^*\|_2} = \nabla f(x^k)^T d^k + \frac{o(\|x^k - x^*\|_2)}{\|x^k - x^*\|_2} \to \nabla f(x^*)^T d \quad \text{as} \quad k \to \infty.$$

Moreover, for $i = 1, \ldots, p$,

$$0 = \frac{h_i(x^k) - h_i(x^*)}{\|x^k - x^*\|_2} = \nabla h_i(x^k)^T d^k + \frac{o(\|x^k - x^*\|_2)}{\|x^k - x^*\|_2} \to \nabla h_i(x^*)^T d \quad \text{as} \quad k \to \infty.$$

and for all $i \in \mathcal{A}(x^*)$,

$$0 \geq \frac{g_i(x^k) - g_i(x^*)}{\|x^k - x^*\|_2} = \nabla g_i(x^k)^T d^k + \frac{o(\|x^k - x^*\|_2)}{\|x^k - x^*\|_2} \to \nabla g_i(x^*)^T d \quad \text{as} \quad k \to \infty.$$

We conclude that $\nabla f(x^*)^T d \leq 0$, $\nabla h_i(x^*)^T d = 0$ for $i = 1, \ldots, p$, and $\nabla g_i(x^*)^T d \leq 0$ for $i \in \mathcal{A}(x^*)$. We further have

$$\begin{aligned} 0 = \nabla_x L(x^*, \lambda^*, \mu^*)^T d &= \nabla f(x^*)^T d + \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla g_i(x^*)^T d + \sum_{i=1}^{p} \mu_i^* \nabla h_i(x^*)^T d \\ &= \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla g_i(x^*)^T d \leq 0. \end{aligned}$$

We obtain $\nabla g_i(x^*)^T d = 0$ for all $i \in \mathcal{A}(x^*)$ with $\lambda_i^* > 0$. Hence $d \in T_+(g, h, x^*, \lambda^*)$.

Using $\lambda^* \geq 0$ and $g(x^k) \leq 0$, we obtain

$$L(x^k, \lambda^*, \mu^*) = f(x^k) + \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* g_i(x^k) \leq f(x^k) \leq f(x^*) = L(x^*, \lambda^*, \mu^*).$$

Combined with $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$, the definition of d^k, and a second-order Taylor's expansion, we obtain

$$0 \geq \frac{L(x^k, \lambda^*, \mu^*) - L(x^*, \lambda^*, \mu^*)}{\|x^k - x^*\|_2^2} = \frac{1}{2} (d^k)^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) d^k + \frac{o(\|x^k - x^*\|_2^2)}{\|x^k - x^*\|_2^2}.$$

Taking limits as $k \to \infty$, we obtain $d^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) d \leq 0$. Since $d \in T_+(g, h, x^*, \lambda^*)$, we have obtained a contradiction to a hypothesis. $\square$

4.4 Exercises

Exercise 4.1.
Consider the inequality constrained problem depicted in Figure 1. Is x^* a KKT point?

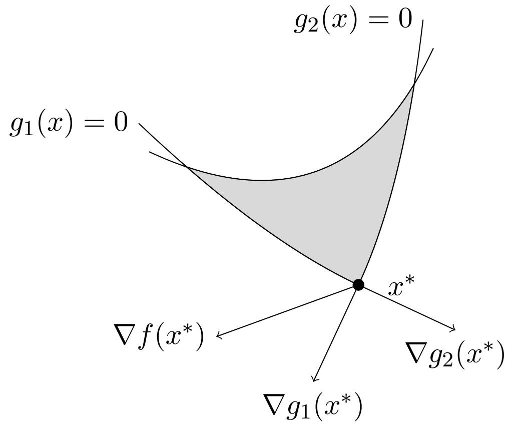

Figure 4.7: The gray region is the feasible set.

Exercise 4.2.
We consider

$$\min_{x \in \mathbb{R}^2} f(x) \qquad \text{s.t.} \qquad h(x) = 0, \tag{P$_1$}$$

where $f : \mathbb{R}^2 \to \mathbb{R}$ and $h : \mathbb{R}^2 \to \mathbb{R}$ are defined by $f(x) := -x_1^4 x_2$ and $h(x) := -(x_1 + 2)^2 + x_2$.

Define $\bar{x} := (-2, 0) \in \mathbb{R}^2$ and $\bar{\mu} := 16$.

1. Show that $(\bar{x}, \bar{\mu})$ is a KKT tuple of problem (P$_1$).
2. Show that $\bar{x}$ is a regular point of (P$_1$)
3. Show that $T_+(h, \bar{x}, \bar{\mu}) = \{\, d \in \mathbb{R}^2 : d_1 \in \mathbb{R},\, d_2 = 0 \,\}$.
4. Define $d := (1, 0)$. Show that $d^T \nabla_{xx} L(\bar{x}, \bar{\mu}) d = -32$, where L is the Lagrangian function corresponding to problem (P$_1$).
5. Show that $\bar{x}$ is not a local solution to problem (P$_1$).

Exercise 4.3.
We consider

$$\min_{x \in \mathbb{R}^3} f(x) \quad \text{s.t.} \quad g_1(x) \le 0, \quad g_2(x) \le 0, \quad h(x) = 0, \tag{P$_2$}$$

where $f(x) := x_1 + x_2 - 2x_3$, $g_1(x) := (1/2)x_1^2 - x_2$, $g_2(x) := \exp(x_1 - 1) - x_1$, and $h(x) := x_3^2 - x_1 + 3/4 = 0$.

We define $\bar{x} := (1, 1/2, 1/2) \in \mathbb{R}^3$, $\bar{\lambda} := (1, 0) \in \mathbb{R}^2$, and $\bar{\mu} := 2$.

1. Show that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple of (P_2).
2. Prove that $T_+(g, h, \bar{x}, \bar{\lambda}) = \{\, (t, t, t) : t \in \mathbb{R} \,\}$.
3. Show that

$$\nabla_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu}) = \begin{bmatrix} 1 & & \\ & 0 & \\ & & 4 \end{bmatrix}.$$

4. Show that $\bar{x}$ is a local solution to (P_2).

Exercise 4.4.
We consider the optimization problem

$$\min_{x \in \mathbb{R}^2} f(x) \quad \text{s.t.} \quad g_1(x) \leq 0, \quad g_2(x) \leq 0, \quad g_3(x) \leq 0, \tag{P$_3$}$$

where $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by $f(x) := x_1^3 x_2$ and $g : \mathbb{R}^2 \to \mathbb{R}$, $i = 1, 2, 3$ are given by

$$g_1(x) := x_1^2 + x_2^2 - 4, \quad g_2(x) := -(x_1 + 2)^2 + x_2, \quad g_3(x) := x_1 - \exp(-x_2).$$

We define $\bar{x} := (-2, 0) \in \mathbb{R}^2$ and $\bar{\lambda} := (0, 8, 0) \in \mathbb{R}^3$

1. Show that $\bar{x}$ is a KKT point of problem (P_3).
2. Show that $\bar{x}$ is a regular point of (P_3).
3. Show that $\bar{x}$ is not a local solution of problem (P_3).

Exercise 4.5 (Uniqueness of Lagrange multipliers under regularity)**.**
We consider

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_i(x) \leq 0,\ i = 1, \ldots, m \quad h_j(x) = 0,\ j = 1, \ldots, p, \tag{P$_4$}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, and $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, p$, are continuously differentiable.

Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and $(\bar{x}, \widehat{\lambda}, \widehat{\mu})$ be KKT triples of (P_4). Suppose that $\bar{x}$ is *regular*. Show that $(\bar{\lambda}, \bar{\mu}) = (\widehat{\lambda}, \widehat{\mu})$.

Exercise 4.6 (KKT conditions for the Celis–Dennis–Tapia problem).
We consider

$$\min_{x\in\mathbb{R}^n} x^T Hx + 2b^T x \quad \text{s.t.} \quad \|x\|_2^2 - \Delta^2 \leq 0, \quad \|A^T x + c\|_2^2 - \xi^2 \leq 0, \tag{P$_5$}$$

where $\Delta > 0$, $\xi \geq 0$, $H \in \mathbb{R}^{n\times n}$ is symmetric, $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}^m$.

Derive the KKT conditions for (P$_5$).

(You only need to state the KKT conditions for (P$_5$) and do not need to compute its KKT points. The KKT conditions for (P$_5$) generally lack closed-form solutions. As a consequence, the computation of KKT points requires numerical computations.)

Exercise 4.7.
We consider the inequality constrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_i(x) \leq 0, \; i = 1, \ldots, m. \tag{P$_6$}$$

Let x^* be a local solution to (P$_6$) and let f and g_i, $i = 1, \ldots, m$, be continuously differentiable on $\mathbb{R}^n$. Suppose that the inequality constraints g_i, $i = 1, \ldots, m$, are concave.

1. Show that x^* is a local solution to the linearized optimization problem

$$\min_{x\in\mathbb{R}^n} f(x^*) + \nabla f(x^*)^T (x - x^*) \quad \text{s.t.} \quad g_i(x^*) + \nabla g_i(x^*)^T (x - x^*) \leq 0, \; i = 1, \ldots, m.$$

2. Is x^* a KKT point of (P$_6$)?

Exercise 4.8 (Second-order sufficient optimality conditions, quadratic growth condition, and strong metric subregularity for unconstrained optimization).
Let $U \subset \mathbb{R}^n$ be nonempty and open, let $f : U \to \mathbb{R}$ be twice continuously differentiable, and let $x^* \in U$.

1. Show that if there exists $\alpha > 0$ such that

$$f(x) - f(x^*) \geq (\alpha/2)\|x - x^*\|_2^2 \quad \text{for all} \quad x \in U, \tag{4.10}$$

then x^* is a strict local minimizer of f over U, $\nabla f(x^*) = 0$, and $d^T \nabla^2 f(x^*) d \geq \alpha\|d\|_2^2$ for all $d \in \mathbb{R}^n$.

2. Show that if there exists $\alpha > 0$ such that

$$\nabla f(x^*) = 0 \quad \text{and} \quad d^T \nabla^2 f(x^*) d \geq \alpha\|d\|_2^2 \quad \text{for all} \quad d \in \mathbb{R}^n, \tag{4.11}$$

then there exists $\varepsilon > 0$ such that

$$f(x) - f(x^*) \geq (\alpha/4)\|x - x^*\|_2^2 \quad \text{for all} \quad x \in U \cap B_\varepsilon(x^*).$$

3. Show that if (4.11) holds for some $\alpha > 0$, then there exists $\varepsilon > 0$ such that

$$(\alpha/2)\|x - x^*\|_2 \leq \|\nabla f(x)\|_2 \quad \text{for all} \quad x \in U \cap B_\varepsilon(x^*). \tag{4.12}$$

Exercise 4.9 (Second-Order Sufficient Condition and Quadratic Growth).
Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ be twice continuously differentiable. We consider

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \tag{P$_7$}$$

Let $\bar{x}$ be a KKT point of (P$_7$) with multipliers $\bar{\lambda} \in \mathbb{R}^m$ and $\bar{\mu} \in \mathbb{R}^p$. We denote by X the feasible set of (P$_7$). Suppose that the second-order sufficient conditions stated in Theorem 4.13 hold true. Show that there exist $\alpha > 0$ and $\delta > 0$ such that

$$f(x) - f(\bar{x}) \geq \alpha\|x - \bar{x}\|_2^2 \quad \text{for all} \quad x \in X \quad \text{with} \quad \|x - \bar{x}\|_2 < \delta.$$

Hint: Adapt the proof of Theorem 4.13.

Exercise 4.10 (LICQ and Slack Variables).
Let $p = 0$ and let $\bar{x}$ be a feasible for (P$_7$). Suppose that the LICQ holds at $\bar{x}$. We consider

$$\min_{(x,s) \in \mathbb{R}^n \times \mathbb{R}^m} f(x) \quad \text{s.t.} \quad g(x) + s = 0, \quad -s \leq 0. \tag{P$_8$}$$

We define $\bar{s} := -g(\bar{x})$.

1. Show that $(\bar{x}, \bar{s})$ is feasible for (P$_8$).
2. Prove that the LICQ holds at $(\bar{x}, \bar{s})$ for (P$_8$).

5 Optimization Methods for Unconstrained Optimization

We consider the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{5.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable.

We study iterative optimization methods that compute (approximate) critical points of (5.1). A point $\bar{x} \in \mathbb{R}^n$ is called *a critical point* or a *stationary point* of f if $\nabla f(\bar{x}) = 0$. A local solution to (5.1) is a critical point of (5.1) (see Proposition 4.6).

5.1 Gradient descent method

The section is devoted to the oldest and most widely known method for unconstrained minimization problems, the gradient descent (method).

Let us provide a "derivation" of the gradient descent method. A vector s is called a *descent direction* of f at $x \in \mathbb{R}^n$ if $\nabla f(x)^T s < 0$. Suppose that $x \in \mathbb{R}^n$ is not a critical point of f, that is, $\nabla f(x) \neq 0$. Then $g \coloneqq -\nabla f(x)$ is a *descent* direction of f at x:

$$\lim_{\gamma \to 0} \frac{f(x - \gamma \nabla f(x)) - f(x)}{\gamma} = -\|\nabla f(x)\|_2^2 < 0.$$

Moreover, the descent direction $g = -\nabla f(x)$ is the *best* among the descent directions h with the same norm as g: For each $h \in \mathbb{R}^n$ with $\|h\|_2 = \|g\|_2$, we have

$$\lim_{\gamma \to 0+} \frac{f(x - \gamma \nabla f(x)) - f(x)}{\gamma} = \nabla f(x)^T h$$

and using the Cauchy–Schwartz inequality, we obtain

$$\nabla f(x)^T h \geq -\|h\|_2 \|\nabla f(x)\|_2 = -\|\nabla f(x)\|_2^2$$

and equality holds if and only if $h = g = -\nabla f(x)$. In particular, if x is not a stationary point of f, then the points $x - \gamma \nabla f(x)$ decrease f for all sufficiently small $\gamma > 0$.

Algorithm 5.1 (Generic gradient descent method)**.**

0. Choose initial point/starting point $x^0 \in \mathbb{R}^n$.

For $k = 0, 1, 2, \ldots$

1. Compute a step size $\gamma_k \geq 0$ and define $x^{k+1} = x^k - \gamma_k \nabla f(x^k)$.

The key issue in gradient descent is the choice of the step size γ_k. As a basic requirement, we could impose is the following: Compute a step size $\gamma_k \geq 0$ that satisfies

$$\nabla f(x^k) \neq 0 \quad \text{implies} \quad f(x^{k+1}) < f(x^k). \tag{5.2}$$

In other words, a step size should ensure that the objective function value decreases as long as the iterates are not critical points.

The gradient step $s^k = -\nabla f(x^k)$ is the solution to

$$\min_{s \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T s + (1/2)\|s\|_2^2.$$

5.1.1 Step Size Selection

We discuss standard step size rules.

- Constant step sizes: The choice $\gamma_k = \gamma > 0$ for all iterations $k \in \mathbb{N}$ is made. Generally, this step size rule may violate (5.2). Under certain conditions on f and on γ, we can ensure (5.2). See Exercise 5.1. Exercise 5.12 shows that improper choices of γ can prevent the gradient method from converging to stationary points.

- Diminishing step sizes: Choose $\gamma_k > 0$ with $\gamma_k \to 0$ as $k \to \infty$ and

  $$\sum_{k=1}^{\infty} \gamma_k = +\infty.$$

  Sometimes the additional requirement

  $$\sum_{k=1}^{\infty} \gamma_k^2 < +\infty$$

  is imposed.

  Exercise 5.15 shows that the use of summable, positive step sizes, that is, $\gamma_k > 0$ and $\sum_{k=1}^{\infty} \gamma_k < +\infty$, may prevent the gradient method from computing approximate stationary points. We provide a simple example illustrating this point.

  Example 5.2. *Let us apply the gradient method to $f(x) := (1/2)x^2$ which has the unique minimizer $x^* = 0$. The derivative f' is Lipschitz continuous with Lipschitz constant* 1. *We choose $\gamma_k := 1/[(k+1)^2\pi^2]$. We have $\sum_{k=1}^{\infty} \gamma_k < +\infty$. Let $x^{k+1} = x^k - \gamma_k x^k$ with $x^0 := 1$. Using Exercise* 5.2 *(or Lemma* 5.5*), we can show that $f(x^{k+1}) < f(x^k)$ for all $k \in \mathbb{N}$. Moreover*

  $$x^{k+1} = (1-\gamma_k)x^k = \left(1 - \frac{1}{(k+1)^2\pi^2}\right)x^k = \prod_{j=0}^{k}\left(1 - \frac{1}{(j+1)^2\pi^2}\right) \to \sin(1) \quad \textit{as} \quad k \to \infty.$$

In particular, (x^k) does not converge to 0 as $k \to \infty$. This example shows that the use of summable, positive step sizes may prevent the gradient method from generating sequences converging to stationary points. Moreover even if the gradient method generates a sequence (x^k) with $f(x^{k+1}) < f(x^k)$ for all $k \in \mathbb{N}$, we can have $\|\nabla f(x^k)\|_2 \geq c$ for all $k \in \mathbb{N}$ and some $c > 0$.

- Minimization rule: The step size γ_k is computed according to

$$\gamma_k = \operatorname{argmin}_{\gamma \geq 0} f(x^k - \gamma \nabla f(x^k)).$$

The step size γ_k is well-defined if $\gamma \to f(x^k - \gamma \nabla f(x^k))$ is bounded below on $[0, \infty)$. For general f, it may be difficult to compute γ_k according to the minimization rule.

- Armijo rule: Compute $\gamma_k > 0$ such that

$$\begin{aligned} f(x^k - \gamma_k \nabla f(x^k)) &\leq f(x^k) - \varepsilon \gamma_k \|\nabla f(x^k)\|_2^2, \\ f(x^k - \eta \gamma_k \nabla f(x^k)) &\geq f(x^k) - \varepsilon \eta \gamma_k \|\nabla f(x^k)\|_2^2, \end{aligned} \tag{5.3}$$

where $\varepsilon \in (0, 1)$ (e.g. $\varepsilon = 1/5$) and $\eta > 1$ (e.g. $\eta = 2$ or $\eta = 10$) are fixed parameters. As opposed to the minimization rule, the Armijo rule is implementable if f is continuously differentiable and bounded from below. The first condition in (5.3) ensures decrease of f provided that $\nabla f(x^k) \neq 0$. Second condition ensures γ_k is "not too small": if we multiply γ_k by η, then first condition is violated. The step sizes computed via the Armijo rule satisfy (5.2).

The Armijo rule has the following geometric interpretation. Let us fix $x^k \in \mathbb{R}^n$ with $\nabla f(x^k) \neq 0$. We consider the function

$$\phi(\gamma) \coloneqq f(x^k - \gamma \nabla f(x^k)) - f(x^k).$$

We have $\phi'(0) = -\nabla f(x^k)^T \nabla f(x^k)$ and $\phi(0) = 0$. Hence the conditions in (5.3) can be expressed as

$$\phi(\gamma_k) \leq \varepsilon \gamma_k \phi'(0) \quad \text{and} \quad \phi(\eta \gamma_k) \geq \varepsilon \eta \gamma_k \phi'(0). \tag{5.4}$$

Figure 5.1 provides a graphical illustration.

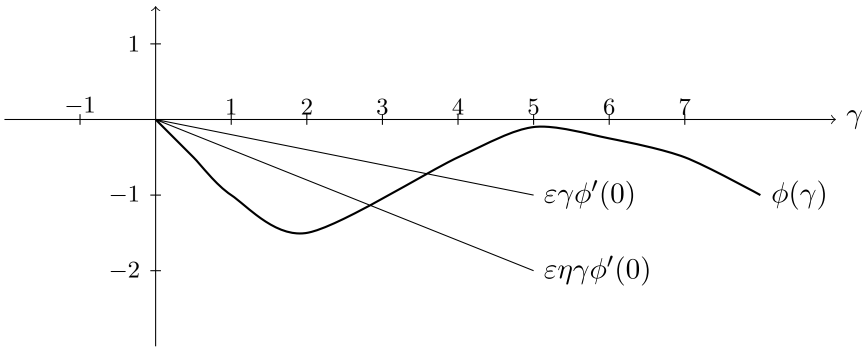

Figure 5.1: We define $\phi(\gamma) \coloneqq f(x^k - \gamma\nabla f(x^k)) - f(x^k)$. We have $\phi(0) = 0$. The figure depicts the graph of ϕ, $\varepsilon\gamma\phi'(0)$, and $\varepsilon\eta\gamma\phi'(0)$ as a function of $\gamma \geq 0$. Here $\phi'(0) = -1$, $\varepsilon = 1/5$ and $\eta = 2$. If $\gamma_k = 5/2$, then the conditions in (5.4) hold true. If $\gamma_k = 1$, then the second condition in (5.4) is violated.

Let f be bounded from below, that is, $f(x) \geq f^*$ for all $x \in \mathbb{R}^n$ and for some $f^* \in \mathbb{R}$. Let us show that for each x^k with $\nabla f(x^k) \neq 0$, there exists a step size $\gamma_k > 0$ fulfilling (5.3). Let us define

$$\phi(\gamma) \coloneqq f(x^k - \gamma\nabla f(x^k)) - f(x^k).$$

We have

$$\phi'(0) = -\|\nabla f(x^k)\|_2^2 \quad \text{and} \quad \phi(0) = 0.$$

So the conditions in (5.3) can be written as

$$\phi(\gamma_k) \leq \phi(0) + \varepsilon\gamma_k\phi'(0), \quad \text{and} \quad \phi(\eta\gamma_k) \geq \phi(0) + \varepsilon\eta\gamma_k\phi'(0).$$

We show that for each sufficiently small $\gamma_k > 0$ the first condition hold true. Since $\nabla f(x^k) \neq 0$, we have $\phi'(0) < 0$. Moreover,

$$\lim_{\gamma \to +0} \frac{\phi(\gamma) - \phi(0)}{\gamma} = \phi'(0).$$

Combined with $\varepsilon \in (0, 1)$, we find that for all sufficiently small $\gamma > 0$,

$$\frac{\phi(\gamma) - \phi(0)}{\gamma} \leq \varepsilon\phi'(0).$$

Multiplying by $\gamma > 0$, we obtain $\phi(\gamma) - \phi(0) \leq \gamma\varepsilon\phi'(0)$.

We must yet show that $\phi(\gamma) - \phi(0) \leq \gamma\varepsilon\phi'(0)$ cannot be valid for all sufficiently large values of $\gamma > 0$. Indeed, we have $\gamma\varepsilon\phi'(0) \to -\infty$ as $\gamma \to \infty$ because $\phi'(0) < 0$ and $\varepsilon > 0$. However, the function $\phi(\gamma) - \phi(0)$ is bounded from below, as f is bounded from below.

- Goldstein test. Let us define
$$\phi(\gamma) := f(x^k - \gamma\nabla f(x^k))$$
and let $\varepsilon \in (0, 1/2)$. A step size $\gamma_k > 0$ satisfies the Goldstein test if
$$\phi(0) + (1-\varepsilon)\gamma\phi'(0) \le \phi(\gamma) \le \phi(0) + \varepsilon\gamma\phi'(0).$$

The minimization rule and the Armijo rule are *line search methods.* Line search methods are aimed at approximately minimizing
$$\gamma \mapsto f(x^k + \gamma s^k)$$
over $\gamma \ge 0$ (or more generally a closed interval). In our case, we have $s^k = -\nabla f(x^k)$. Further line search methods include the Goldstein rule and the Wolfe line search, for example. Line search schemes that successfully reduce γ until a termination criterion is reached are called *backtracking line search methods.*

The Armijo rule as introduced in (5.3) is a special case of the following conditions applied to the gradient step $s^k = -\nabla f(x^k)$. For a step $s^k \in \mathbb{R}^n$, the Armijo rule is: Compute $\gamma_k > 0$ such that
$$\begin{aligned} f(x^k + \gamma_k s^k) &\le f(x^k) + \varepsilon\gamma_k \nabla f(x^k)^T s^k, \\ f(x^k + \eta\gamma_k s^k) &\ge f(x^k) + \varepsilon\eta\gamma_k \nabla f(x^k)^T s^k, \end{aligned} \tag{5.5}$$
where $\varepsilon \in (0,1)$ and $\eta > 1$ are fixed parameters. Whenever $\nabla f(x^k)^T s^k < 0$, a step size $\gamma_k > 0$ satisfying the conditions in (5.5) can be computed. The verification of this statement is similar to that presented above for the case $s^k = -\nabla f(x^k)$.

A step satisfying the Armijo conditions (5.5) can be computed using backtracking line search, provided that $\nabla f(x^k)^T s^k < 0$:

Compute the largest number $\gamma_k \in \{1, (1/\eta), (1/\eta)^2, \ldots\}$ such that
$$f(x^k + \gamma_k s^k) \le f(x^k) + \varepsilon\gamma_k \nabla f(x^k)^T s^k. \tag{5.6}$$
This particular scheme to satisfy the Armijo conditions presupposes that $\gamma_k = 1$ might be a "good" step size. Figure 5.2 provides a graphical illustration.

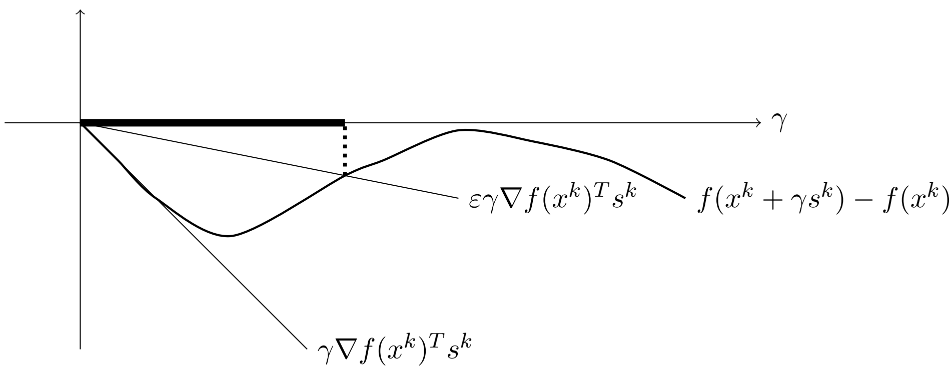

Figure 5.2: Backtracking line search by the Armijo rule (5.6). The thick line corresponds to the interval of step sizes γ_k fulfilling the condition in (5.6).

5.1.2 Global convergence

We demonstrate the (asymptotic) convergence of the gradient descent, Algorithm 5.1, if the Armijo step size rule is used. More precisely, we show that each accumulation point of the sequence generated by the gradient descent is a critical point of f. This convergence result does *not* imply that the accumulation point is a global or even local solution to (5.1).

Theorem 5.3 (Global convergence of gradient descent with Armijo rule)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and bounded from below. We consider Algorithm* 5.1 *with Armijo step size rule. Then the algorithm generates a sequence (x^k) such that*

1. *for all $k \in \mathbb{N}$, $f(x^{k+1}) < f(x^k)$ with $\nabla f(x^k) \neq 0$, and $f(x^{k+1}) = f(x^k)$ otherwise, and*

2. *each accumulation point of (x^k) is a stationary point of f.*

Proof. We show that the algorithm computes the iterates $x^1, x^2, \ldots$. In other words, we show that the method is well-defined. We establish this assertion using induction. If $\nabla f(x^0) = 0$, then any $\gamma_1 > 0$ satisfies (5.3) and we have $x^1 = x^0$. If $\nabla f(x^0) \neq 0$, then the computations performed in Section 5.1.1 show that a step size $\gamma_1 > 0$ can be computed such that the conditions in (5.3) hold true. Using (5.3), we obtain

$$f(x^1) - f(x^0) \leq -\varepsilon\gamma_1 \|\nabla f(x^0)\|_2^2 < 0.$$

Now suppose that the algorithm has computed x^k. Using arguments similar to those above, we can show that the algorithm computes the next iterate x^{k+1}. Hence, we have shown that the algorithm generates a sequence (x^k) and that $f(x^{k+1}) < f(x^k)$ for all $k \in \mathbb{N}$ with $\nabla f(x^k) \neq 0$.

Now let $\bar{x}$ be an accumulation point of (x^k) and let (x^{k_ℓ}) be a subsequence of (x^k) converging to $\bar{x}$ as $\ell \to \infty$.

Since the sequence $(f(x^k))$ is nonincreasing and f is bounded from below, we find that $f(x^k) \to f(\bar{x})$ as $k \to \infty$ (This follows from the "monotone convergence theorem"[1] and the continuity of f.) In particular, we have $f(x^k) - f(x^{k+1}) \to 0$ as $k \to \infty$. Using (5.3), we find that

$$\varepsilon\gamma_k \|\nabla f(x^k)\|_2^2 \leq f(x^k) - f(x^{k+1}).$$

Combined with $f(x^k) - f(x^{k+1}) \to 0$ as $k \to \infty$, we find that

$$\gamma_k \|\nabla f(x^k)\|_2^2 \to 0 \quad \text{as} \quad k \to \infty. \tag{5.7}$$

[1] "Monotone convergence theorem:" If a sequence of real numbers is nonincreasing and bounded from below by a finite number, then it converges. In other words, if $\bar{a} \in \mathbb{R}$, $(a_k) \subset \mathbb{R}$ is a sequence, and $a_k \geq \bar{a}$ for all $k \in \mathbb{N}$, then $\lim_{k\to\infty} a_k$ exists.

To establish the theorem, we must show that $\nabla f(\bar{x}) = 0$. Suppose that $\nabla f(\bar{x}) \neq 0$. Combined with the continuity of ∇f and $x^{k_\ell} \to \bar{x}$ as $\ell \to \infty$, we find that for all sufficiently large $\ell \in \mathbb{N}$,

$$\|\nabla f(x^{k_\ell})\|_2 \geq (1/2)\|\nabla f(\bar{x})\| > 0.$$

Using (5.7), we have $\gamma_{k_\ell} \to 0$ as $\ell \to \infty$. Let us define $t_\ell := \eta\gamma_{k_\ell}$ and $s^k := -\nabla f(x^k)$. The sequence (t_ℓ) converges to zero as $\ell \to \infty$.

Using (5.3), we obtain for all ℓ,

$$f(x^{k_\ell} + \eta\gamma_{k_\ell} s^{k_\ell}) - f(x^{k_\ell}) \geq -\varepsilon\eta\gamma_{k_\ell}\|\nabla f(x^{k_\ell})\|_2^2. \tag{5.8}$$

Dividing (5.8) by $\eta\gamma_{k_\ell}$, using $t_\ell = \eta\gamma_{k_\ell}$, and using the continuity of ∇f, we obtain

$$\liminf_{\ell\to\infty} \frac{f(x^{k_\ell} + t_\ell s^{k_\ell}) - f(x^{k_\ell})}{t_\ell} \geq -\varepsilon\|\nabla f(\bar{x})\|_2^2.$$

Next we show that

$$\lim_{\ell\to\infty} \frac{f(x^{k_\ell} + t_\ell s^{k_\ell}) - f(x^{k_\ell})}{t_\ell} = -\|\nabla f(\bar{x})\|_2^2.$$

The mean value theorem ensures the existence of numbers $\tau_\ell \in (0, t_\ell)$ such that

$$\frac{f(x^{k_\ell} + t_\ell s^{k_\ell}) - f(x^{k_\ell})}{t_\ell} = \frac{t_\ell \nabla f(x^{k_\ell} + \tau_\ell s^{k_\ell})^T s^{k_\ell}}{t_\ell} = \nabla f(x^{k_\ell} + \tau_\ell s^{k_\ell})^T s^{k_\ell}.$$

Combined with the continuity of ∇f and $s^{k_\ell} = -\nabla f(x^{k_\ell}) \to -\nabla f(\bar{x})$ as $k \to \infty$, we obtain

$$\lim_{\ell\to\infty} \frac{f(x^{k_\ell} + t_\ell s^{k_\ell}) - f(x^{k_\ell})}{t_\ell} = -\|\nabla f(\bar{x})\|_2^2.$$

Hence $-\|\nabla f(\bar{x})\|_2^2 \geq -\varepsilon\|\nabla f(\bar{x})\|_2^2$, yielding $0 \geq (1-\varepsilon)\|\nabla f(\bar{x})\|_2^2$. Since $\varepsilon \in (0,1)$, we obtain $\nabla f(\bar{x}) = 0$. $\square$

The assertions of Theorem 5.3 remain valid if instead of the Armijo rule the minimization rule is used; see Exercise 5.6.

Theorem 5.3 provides conditions sufficient for accumulation points of the sequence (x^k) generated by the gradient descent be stationary points of f. When does the sequence have an accumulation point? If the level set

$$\{x \in \mathbb{R}^n \colon f(x) \leq f(x^0)\}$$

is bounded, then the sequence (x^k) has at least one accumulation point because Theorem 5.3 ensures that the iterates x^k belong to this level set.

Example 5.4. The gradient method as applied to a nonconvex function may not generate a sequence converging to a local minimizer of it. For example, let us consider $f(x) = x^4 + x^3 + 1/4$. We have $f'(x) = 4x^3 + 3x^2 = x^2(4x+3) = 0$ if and only if $x = 0$ or $x = -3/4$. The point $x^* = -3/4$ is the global minimizer of f, while $\bar{x} = 0$ is a saddle point (a critical point that is neither a local minimizer nor local maximizer). If we initialize the gradient descent with $x^0 = 0$, then we have $0 = x^0 = x^1 = x^2 = \cdots$. In particular, x^k does not converge to $x^* = -3/4$. Figure 5.3 provides an illustration.

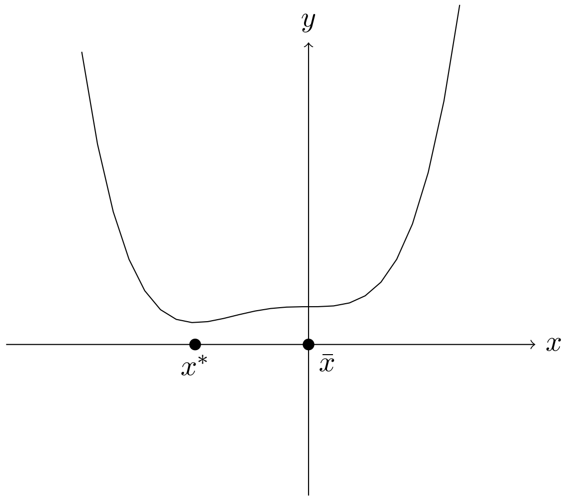

Figure 5.3: Graph of the function $f(x) = x^4 + x^3 + 1/4$. The point x^* is the global minimizer of f, and $\bar{x}$ is a saddle point of f (a critical point that is neither a local minimizer nor local maximizer).

5.1.3 Convergence rates for nonconvex objectives

Theorem 5.3 provides an asymptotic convergence result. We now demonstrate a nonasymptotic convergence statement.

We say that the differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ has a Lipschitz continuous gradient if there exists a constant $L \geq 0$ such that

$$\|\nabla f(y) - \nabla f(x)\|_2 \leq L\|y - x\|_2 \quad \text{for all} \quad x, y \in \mathbb{R}^n.$$

In this case, L is referred to as the Lipschitz constant of ∇f.

The nonasymptotic convergence rates are based on the following lemma.

Lemma 5.5 (Descent lemma). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable, and let ∇f be Lipschitz continuous with Lipschitz constant $L > 0$. Then*

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + (L/2)\|y - x\|_2^2 \quad \textit{for all} \quad x, y \in \mathbb{R}^n. \tag{5.9}$$

Proof. Fix x, $y \in \mathbb{R}^n$. We have

$$f(y) - f(x) = \int_0^1 \nabla f(x + \tau(y - x))^T(y - x)d\tau.$$

Combined with the Lipschitz continuity, we obtain

$$\begin{aligned} f(y) - f(x) - \nabla f(x)^T(y-x) &= \int_0^1 \left[\nabla f(x+\tau(y-x))^T(y-x) - \nabla f(x)^T(y-x)\right] d\tau \\ &\le \int_0^1 \|\nabla f(x+\tau(y-x)) - \nabla f(x)\|_2 \|y-x\|_2 d\tau \\ &\le \|y-x\|_2^2 \int_0^1 L\tau d\tau \\ &= (L/2)\|y-x\|_2^2. \end{aligned}$$

□

Let us minimize the right-hand side in (5.9) over $y \in \mathbb{R}^n$. The right-hand side is a strongly convex quadratic function. For its minimizer y^*, we obtain

$$\nabla f(x) + L(y^* - x) = 0,$$

yielding $y^* = x - (1/L)\nabla f(x)$. Hence

$$\min_{y\in\mathbb{R}^n} \left\{f(x) + \nabla f(x)^T(y-x) + (L/2)\|y-x\|_2^2\right\} = f(x) - \frac{1}{2L}\|\nabla f(x)\|_2^2$$

and

$$f(x - (1/L)\nabla f(x)) \le f(x) - \frac{1}{2L}\|\nabla f(x)\|_2^2.$$

We state global convergence rates for the gradient descent method, Algorithm 5.1.

Theorem 5.6. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and its gradient be Lipschitz continuous with Lipschitz constant L. Let $f^* \in \mathbb{R}$ be the optimal value of* (5.1)*. Then for all $K \in \mathbb{N}$:*

1. *The sequence (x^k) generated by the gradient descent method with minimization rule satisfies*

$$\min_{0\le k<K} \|\nabla f(x^k)\|_2^2 \le \frac{2L}{K} \cdot (f(x^0) - f^*).$$

2. *The sequence (x^k) generated by the gradient descent method with Armijo rule satisfies*

$$\min_{0\le k<K} \|\nabla f(x^k)\|_2^2 \le \frac{L\eta}{2\varepsilon(1-\varepsilon)K} \cdot (f(x^0) - f^*).$$

Proof. 1. The minimization rule computes $\gamma_k \ge 0$ as a solution to $\min_{\gamma\ge 0} f(x^k - \gamma\nabla f(x^k))$. Choosing $y = x^k - \gamma\nabla f(x^k)$ and $x = x^k$ in Lemma 5.5, and minimizing both sides over $\gamma \geq 0$, we obtain

$$\begin{aligned} f(x^{k+1}) &= \min_{\gamma \geq 0} f(x^k - \gamma \nabla f(x^k)) \\ &\leq f(x^k) + \min_{\gamma \geq 0} \{-\gamma \|\nabla f(x^k)\|_2^2 + (L/2)\gamma^2 \|\nabla f(x^k)\|_2^2\} \\ &= f(x^k) - \frac{1}{2L}\|\nabla f(x^k)\|_2^2. \end{aligned}$$

Rearranging terms, we obtain

$$f(x^k) - f(x^{k+1}) \geq \frac{1}{2L}\|\nabla f(x^k)\|_2^2.$$

Summing over $k = 0, \dots, K$, we have

$$\frac{1}{2L}\sum_{k=0}^{K} \|\nabla f(x^k)\|_2^2 \leq \sum_{k=0}^{K} \left[f(x^k) - f(x^{k+1})\right].$$

We deduce

$$\sum_{k=0}^{K} \left[f(x^k) - f(x^{k+1})\right] = f(x^0) - f(x^1) + f(x^1) - f(x^2) + \cdots = f(x^0) - f(x^{K+1})$$

Putting together the pieces, we obtain

$$\frac{1}{2L}\sum_{k=0}^{K} \|\nabla f(x^k)\|_2^2 \leq f(x^0) - f(x^{K+1}).$$

The term on the left-hand side is greater than or equal to $(K+1)/(2L)\min_{0\leq k<K+1} \|\nabla f(x^k)\|_2^2$ and the term on the right-hand side is less than or equal to $f(x^0) - f^*$. We obtain the assertion.

2. We apply Lemma 5.5 to $y = x^k - \gamma\eta\nabla f(x^k)$ and $x = x^k$ to obtain

$$f(x^k - \gamma_k\eta\nabla f(x^k)) \leq f(x^k) - \gamma_k\eta\|\nabla f(x^k)\|_2^2 + (L/2)\eta^2\gamma_k^2\|\nabla f(x^k)\|_2^2.$$

Using the second condition in the Armijo rule (5.3),

$$f(x^k - \gamma_k\eta\nabla f(x^k)) \geq f(x^k) - \varepsilon\eta\gamma_k\|\nabla f(x^k)\|_2^2.$$

Combining these inequalities, we get

$$(1-\varepsilon)\eta\gamma_k\|\nabla f(x^k)\|_2^2 \leq (L/2)\eta^2\gamma_k^2\|\nabla f(x^k)\|_2^2.$$

If $\nabla f(x^k) \neq 0$ and $\gamma_k > 0$, we obtain

$$\gamma_k \geq \frac{2(1-\varepsilon)}{\eta L}. \tag{5.10}$$

Using the first condition in the Armijo rule (5.3), we have

$$f(x^k) - f(x^{k+1}) \geq \varepsilon\gamma_k \|\nabla f(x^k)\|_2^2.$$

Combined with (5.10),

$$f(x^k) - f(x^{k+1}) \geq \frac{2\varepsilon(1-\varepsilon)}{\eta L}\|\nabla f(x^k)\|_2^2.$$

Now, we can proceed as in part one.

□

5.1.4 Convergence rates for convex objectives

Theorem 5.6 provides convergence rates for the decay of $\min_{0\leq k\leq K}\|\nabla f(x^k)\|_2$ when (x^k) is generated by a gradient method with either the minimization or the Armijo step size rule. If we additionally assume that f is convex, then we can establish convergence rates for $f(x^k) - f^*$, where f^* is the optimal value of (5.1).

Theorem 5.7. *We consider the gradient descent method with Armijo line search and parameter $\varepsilon \in (1/2, 1)$. Let f be convex and differentiable. Let x^* be a minimizer of f. Suppose that ∇f is Lipschitz continuous with Lipschitz constant L. Then for every $K \in \mathbb{N}$, we have*

$$f(x^K) - f(x^*) \leq \frac{\eta L\|x^0 - x^*\|_2^2}{4(1-\varepsilon)K}.$$

Proof. Let us define $r_k = \|x^k - x^*\|_2$. We have

$$\begin{aligned} r_{k+1}^2 &= \|x^{k+1} - x^*\|_2^2 \\ &= r_k^2 - 2\gamma_k \nabla f(x^k)^T(x^k - x^*) + \gamma_k^2\|\nabla f(x^k)\|_2^2. \end{aligned} \tag{5.11}$$

Since f is convex, the gradient inequality yields

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) \quad \text{for all} \quad x, y \in \mathbb{R}^n.$$

It follows that

$$\nabla f(x^k)^T(x^k - x^*) \geq f(x^k) - f(x^*).$$

Let us define $\delta_k = f(x^k) - f(x^*) \geq 0$. Using (5.11), we obtain

$$r_{k+1}^2 \leq r_k^2 - \gamma_k\big(2\delta_k - \gamma_k\|\nabla f(x^k)\|_2^2\big).$$

According to (5.3), we have

$$\gamma_k\|\nabla f(x^k)\|_2^2 \leq (1/\varepsilon)\big[f(x^k) - f(x^{k+1})\big] = (1/\varepsilon)\big[\delta_k - \delta_{k+1}\big].$$

Since $\gamma_k > 0$, we get

$$r_{k+1}^2 \le r_k^2 - \gamma_k\big[(2-1/\varepsilon)\delta_k + (1/\varepsilon)\delta_{k+1}\big].$$

Since $\varepsilon > 1/2$ and $\delta_k \ge 0$, the quantify in the brackets on the right-hand side is nonnegative. We know from (5.10)

$$\gamma_k \ge \bar{\gamma} \coloneqq \frac{2(1-\varepsilon)}{\eta L}.$$

We obtain

$$r_{k+1}^2 \le r_k^2 - \bar{\gamma}\big[(2-1/\varepsilon)\delta_k + (1/\varepsilon)\delta_{k+1}\big].$$

Rearranging terms yields

$$\bar{\gamma}\big[(2-1/\varepsilon)\delta_k + (1/\varepsilon)\delta_{k+1}\big] \le r_k^2 - r_{k+1}^2.$$

Theorem 5.3 ensures that $\delta_k \ge \delta_{k+1}$, as $f(x^{k+1}) \le f(x^k)$. Therefore

$$2\bar{\gamma}\delta_{k+1} \le r_k^2 - r_{k+1}^2.$$

Summing this inequality over $k = 0, \dots, K-1$, and using $\delta_K \le \delta_k$ for $k \le K$, we have

$$2K\bar{\gamma}\delta_K \le 2\bar{\gamma}\sum_{k=0}^{K-1}\delta_k \le r_0^2 - r_K^2 \le r_0^2 = \|x^* - x^0\|_2^2.$$

Upon using the definition of $\bar{\gamma}$, we obtain the assertion. □

We now state a result on the performance of the gradient descent method with Armijo line search applied to strongly convex functions. A function f is called *strongly* convex with parameter $\mu > 0$ if $f(x) - (\mu/2)\|x\|_2^2$ is a convex function. For example, if $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite and $b \in \mathbb{R}^n$, then $f(x) = (1/2)x^T A x - b^T x$ is strongly convex with parameter $\mu = \lambda_{\min}(A)$ (minimum eigenvalue of A) and ∇f is Lipschitz continuous with Lipschitz constant $L = \lambda_{\max}(A)$ (maximum eigenvalue).

Proposition 5.8. *Let f be strongly convex with parameter $\mu > 0$, and let f be differentiable and its gradient be Lipschitz continuous with Lipschitz constant $L > 0$. If we apply gradient descent with Armijo line search to f with $\varepsilon \in (1/2, 1)$, and x^* is the minimizer of f, then*

$$\|x^k - x^*\|_2 \le \theta^k \|x^0 - x^*\|_2, \quad \textit{where} \quad \theta = \sqrt{\frac{\kappa_f - (2-1/\varepsilon)(1-\varepsilon)/\eta}{\kappa_f + (1/\varepsilon - 1)/\eta}}, \tag{5.12}$$

where $\kappa_f = L/\mu$ is the condition number *of f. Moreover,*

$$f(x^k) - f(x^*) \le \theta^{2k}\kappa_f\big[f(x^0) - f(x^*)\big].$$

We state Proposition 5.8 without a proof. The scalar θ is contained in $(0,1)$, as $\varepsilon \in (1/2,1)$ is assumed in Proposition 5.8. Moreover, if the condition number κ_f is large, then θ is close to 1. The convergence rate in (5.12) is referred to as *linear*.

Why is the ratio L/μ called condition number? In numerical linear algebra, the condition number of a symmetric positive definite matrix is the ratio of the maximum and the minimum eigenvalue. Let $\mu > 0$ and $L > 0$. If f is twice continuously differentiable, then f is strongly convex with parameter μ and ∇f is Lipschitz continuous with Lipschitz constant L if and only if

$$\mu d^T d \leq d^T \nabla^2 f(x) d \leq L d^T d \quad \text{for all} \quad x \in \mathbb{R}^n, \quad d \in \mathbb{R}^n.$$

This is equivalent to $\mu \leq \lambda_{\min}(\nabla^2 f(x)) \leq \lambda_{\max}(\nabla^2 f(x)) \leq L$ for all $x \in \mathbb{R}^n$. These considerations provide a motivation for the terminology *condition number*.

A result similar to Proposition 5.8 is valid if the minimization rule is used.

Proposition 5.9. *Let f be strongly convex with parameter $\mu > 0$, and let f be differentiable and its gradient be Lipschitz continuous with Lipschitz constant $L > 0$. If we apply gradient descent with minimization rule to f, and x^* is the minimizer of f, then*

$$f(x^k) - f(x^*) \leq \left(\frac{\kappa_f - 1}{\kappa_f + 1}\right)^{2k} \big[f(x^0) - f(x^*)\big].$$

and

$$f(x^k) - f(x^*) \leq \left(\frac{\kappa_f - 1}{\kappa_f + 1}\right)^{2k} \big[f(x^{k-1}) - f(x^*)\big].$$

Here, $\kappa_f = L/\mu$.

We state Proposition 5.9 without a proof. However, we consider an illustrative example.

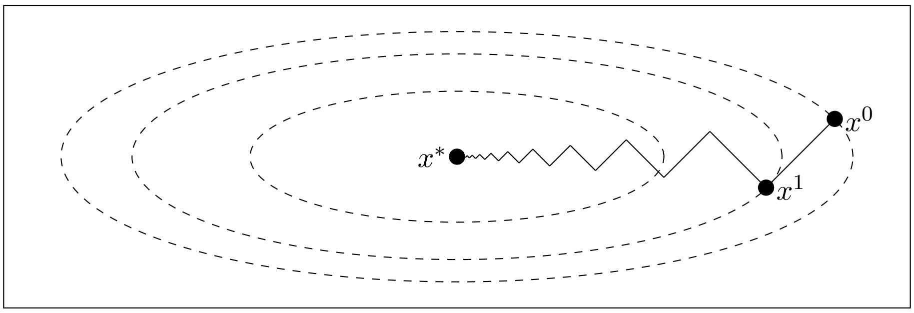

Figure 5.4: Some contour lines of the function $f(x) = (1/2)x_1^2 + (\rho/2)x_2^2$. Contour lines of f are the set of points $x \in \mathbb{R}^2$ with $f(x) = c$ for $c \in \mathbb{R}$. The figure shows iterates of the gradient method with minimization rule and $x^0 = (\gamma, 1)$ with $\gamma = 10$.

Example 5.10. We apply the gradient descent method with minimization rule to the quadratic objective function on $\mathbb{R}^2$,

$$f(x) = \frac{1}{2}\big(x_1^2 + \rho x_2^2\big),$$

where $\rho \geq 1$. The condition number of f is $\kappa_f = \rho/1$. We choose $x^0 = (\rho, 1)$. It can be shown that

$$x_1^k = \rho\Big(\frac{\rho - 1}{\rho + 1}\Big)^k \quad \text{and} \quad x_2^k = \rho\Big(-\frac{\rho - 1}{\rho + 1}\Big)^k$$

and

$$f(x^k) = \left(\frac{\rho - 1}{\rho + 1}\right)^{2k} f(x^0).$$

For this simple example, the converge is linear and the error is reduced by a factor $|(\rho - 1)/(\rho + 1)|^2$ at each iteration. For $\rho = 1$, the exact solution is computed in one iteration. If $\rho \gg 1$, then the convergence is very slow. Moreover, the example shows that the convergence rate in Proposition 5.9 cannot be improved. Figure 5.4 provides an illustration.

5.1.5 Termination

We generally cannot expect that the gradient descent method computes an iterate x^k that is a stationary point of f. In implementations of gradient descent, we may terminate gradient descent if one of the following conditions is satisfied:

$$\|\nabla f(x^k)\|_2 \leq \varepsilon_{\text{atol}} \quad \text{or} \quad \|\nabla f(x^k)\|_2 \leq \varepsilon_{\text{atol}} + \varepsilon_{\text{rtol}}\|\nabla f(x^0)\|_2,$$

where $\varepsilon_{\text{atol}} > 0$ and $\varepsilon_{\text{rtol}} \in (0, 1)$ are tolerances. For the latter termination criterion, we may choose $\varepsilon_{\text{atol}} = 0$. Theorem 5.6 provides us with upper bounds on the number of iterations until $\|\nabla f(x^k)\|_2 \leq \varepsilon_{\text{atol}}$ is satisfied for $\varepsilon_{\text{atol}} > 0$.

5.2 Accelerated gradient descent method

Throughout the section, we consider

$$\min_{x \in X} f(x), \tag{5.13}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is *convex*, and differentiable with L-Lipschitz continuous gradient. Moreover $X \subset \mathbb{R}^n$ is nonempty, closed, and convex.

We discuss only a basic version of the accelerated gradient method (also known as Nesterov's (gradient) method). Comprehensive analyzes of accelerated gradient methods and optimal first-order methods can be found in [13, section 2.2], [11, Chapter 3], and the recent paper [15].

Algorithm 5.11 (basic version of the accelerated gradient descent method)**.**

0. Choose initial point/starting point $x^0 \in X$. Define $\overline{x}^0 := x^0$.

For $k = 1, 2, \dots$

1. Choose $\alpha_k > 0$ and compute $\underline{x}^k = (1-\alpha_k)\overline{x}^{k-1} + \alpha_k x^{k-1}$.
2. Choose $\beta_k > 0$ and compute
$$x^k = \operatorname{argmin}_{y \in X} \left\{ \alpha_k \nabla f(\underline{x}^k)^T y + (\beta_k/2)\|y - x^{k-1}\|_2^2 \right\}.$$
3. Compute $\overline{x}^k = (1-\alpha_k)\overline{x}^{k-1} + \alpha_k x^k$.

We discuss the convergence rate the basic accelerated gradient descent method (see Algorithm 5.11) as applied to smooth convex optimization.

Theorem 5.12. *Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be convex, and differentiable with L-Lipschitz continuous gradient. Moreover $X \subset \mathbb{R}^n$ is nonempty, closed, and convex. Let x^* be a solution to* (5.13)*. Let $(\underline{x}^k, x^k, \overline{x}^k)$ be generated by Algorithm* 5.11 *with $\alpha_k := 2/(k+1)$ and $\beta_k := 4L/[k(k+1)]$. Then for all $k \in \mathbb{N}$,*
$$f(\overline{x}^k) - f(x^*) \le \frac{2L}{k(k+1)}\|x^0 - x^*\|_2^2.$$

Proof. See Theorem 3.6 in [11]. □

5.3 Conjugate gradient method

The conjugate gradient (CG) method was originally designed to solve the quadratic, strongly convex optimization problem
$$\min_{x \in \mathbb{R}^n} f(x), \quad \text{where} \quad f(x) := (1/2)x^T H x - b^T x, \tag{5.14}$$
where $H \in \mathbb{R}^{n \times n}$ is symmetric positive definite and $b \in \mathbb{R}^n$. The solution x^* to (5.14) is given by $x^* = H^{-1}b$.

Let us recall a few facts about gradient descent with minimization rule. When applying gradient descent with minimization rule to the objective function in (5.14), then the first iterate x^1 is computed as $x^0 - \gamma_0 \nabla f(x^0)$, where γ_0 solves
$$\gamma_0 = \operatorname{argmin}_{\gamma \ge 0} f(x^0 - \gamma_0 \nabla f(x^0)).$$

Since f is quadratic and strongly convex and any vector $x \in x^0 + \operatorname{Lin}(\nabla f(x^0))$ is of the form $x = x^0 - \gamma \nabla f(x^0)$ for some $\gamma \in \mathbb{R}$, we may also write
$$x^1 = \operatorname{argmin}_{x \in x^0 + \operatorname{Lin}(\nabla f(x^0))} f(x).$$

Similarly, the kth iterate of the gradient descent using minimization rule is given by

$$x^k = \operatorname{argmin}_{x \in x^{k-1} + \mathrm{Lin}(\nabla f(x^{k-1}))} f(x).$$

The conjugate gradient method uses a different approach. Instead of computing x^k as a minimizer of f over the affine space $x^{k-1} + \mathrm{Lin}(\nabla f(x^{k-1}))$, it computes x^k by minimizing f over a certain linear subspace shifted by x^0. These linear subspaces are called Krylov subspaces.

Let us formalize the conjugate gradient method. Let $x^0 \in \mathbb{R}^n$ be a starting point. We associate with x^0 the *Krylov vectors*

$$g_0 \coloneqq Hx^0 - b, \quad Hg_0, \quad H^2 g_0, \quad \ldots$$

and the *Krylov subspaces*

$$L_0 \coloneqq \{0\}, \quad L_k \coloneqq \mathrm{Lin}(g_0, Hg_0, H^2 g_0, \ldots, H^{k-1} g_0), \quad k = 1, 2, \ldots.$$

Hence L_k is the linear span of the first k Krylov vectors. Since $Hx^* = b$, we can write $g_0 = H(x^0 - x^*)$, $Hg_0 = H^2(x^0 - x^*)$, $H^2 g_0 = H^3(x^0 - x^*)$, …, $H^{k-1} g_0 = H^k(x^0 - x^*)$. We obtain

$$L_k = \mathrm{Lin}(H(x^0 - x^*), H^2(x^0 - x^*), \ldots H^k(x^0 - x^*)).$$

Moreover, we have $L_{k+1} = \mathrm{Lin}(L_k, H^{k+1}(x^0 - x^*))$ and

$$L_1 \subset L_2 \subset L_3 \subset \cdots.$$

Let K be the first value of k such that the first k Krylov vectors are linearly independent. Then the inclusion $L_k \subset L_{k+1}$ is strict when $K < k-1$ and is an equality if $K \geq k-1$.

The classical CG method computes x^k as

$$x^k = \operatorname{argmin}_{x \in x^0 + L_k} f(x).$$

We terminate the CG method if $\nabla f(x^k) = 0$. Below we analyze properties of the iterates x^k as well as the subspaces L_k. These considerations will allow us to state an explicit form of the CG method that is implementable.

We say that a vector $g \in \mathbb{R}^n$ is orthogonal to a linear subspace L if $g^T x = 0$ for all $x \in L$. The vectors d_1, d_2 are said to be orthogonal if $d_1^T d_2 = 0$. If $d_0, \ldots, d_k$ are vectors in $\mathbb{R}^n$ and A is symmetric positive definite, then we say that they are A-conjugate if $d_j^T A d_\ell = 0$ for $0 \leq j < \ell \leq k$. The CG method generates H-conjugate vectors as we show in Lemma 5.14.

Lemma 5.13 provides us with an equivalent representation of the Krylov subspaces and shows that the gradients $g_k = \nabla f(x^k)$ are orthogonal to each other.

Lemma 5.13. *If the CG method does not terminate at step k, then the gradients $g_0, \ldots, g_{k-1}$ of f at the points $x^0, \ldots, x^{k-1}$ are nonzero and*

$$L_k = \mathrm{Lin}(g_0, g_1, \ldots, g_{k-1}).$$

Moreover, we have $g_k^T g_\ell = 0$ for $0 \leq \ell \leq k-1$ and g_k is orthogonal to L_k.

Proof. For $k = 1$, then $\nabla f(x^0) = Hx^0 - b \neq 0$ and we have $L_1 = \mathrm{Lin}(g_0)$. Since $x^1 \in x^0 + L_1$, we have $x^1 = x^0 + \mu^* g_0$ for some $\mu^* \neq 0$. Since $x^1 = \mathrm{argmin}_{x \in x^0 + L_1} f(x)$, μ^* minimizes $\phi(\mu) = f(x^0 + \mu g_0)$. We obtain $0 = \phi'(\mu^*) = \nabla f(x^1)^T g_0 = g_1^T g_0 = 0$.

Suppose that the assertion is true for some k. Since $x^k \in x^0 + L_k$, we have $x^k = x^0 + \sum_{\ell=1}^k \lambda_\ell H^\ell (x^0 - x^*)$ for some $\lambda_\ell \in \mathbb{R}$. We compute

$$\begin{aligned} g_k = \nabla f(x^k) &= H(x^k - x^*) \\ &= H(x^0 - x^*) + \sum_{\ell=1}^{k} \lambda_\ell H^{\ell+1}(x^0 - x^*) \\ &= \underbrace{H(x^0 - x^*) + \sum_{\ell=1}^{k-1} \lambda_\ell H^{\ell+1}(x^0 - x^*)}_{=y^k} + \lambda_k H^{k+1}(x^0 - x^*) \\ &= y^k + \lambda_k H^{k+1}(x^0 - x^*). \end{aligned}$$

We have $y^k \in L_k$. Hence

$$\mathrm{Lin}(L_k, g_k) \subset \mathrm{Lin}(L_k, H^{k+1}(x^0 - x^*)) = L_{k+1}. \tag{5.15}$$

The induction hypothesis ensures

$$\mathrm{Lin}(L_k, \nabla f(x^k)) = \mathrm{Lin}(g_0, \ldots, g_{k-1}).$$

Combined with $x^k = \mathrm{argmin}_{x \in x^0 + L_k} f(x)$, we find that

$$x^k = x^0 + \sum_{j=1}^{k} \mu_j^* g_{j-1}.$$

for some $\mu_j^* \in \mathbb{R}$. Let us consider the function

$$\phi(\mu) = f\left(x^0 + \sum_{j=1}^{k} \mu_j g_{j-1}\right).$$

The vector μ^* is a minimizer of ϕ. Hence we have $\nabla \phi(\mu^*) = 0$. We obtain for $1 \leq j \leq k$,

$$\frac{\partial \phi(\mu^*)}{\partial \mu_j} = g_{j-1}^T g_k = 0.$$

Therefore, the vectors $g_0, \ldots, g_{k-1}$ are orthogonal to $g_k \neq 0$. Consequently, the dimension of $\mathrm{Lin}(L_k, g_k)$ is equal to $k+1$. Combined with (5.15), we find that $\mathrm{Lin}(L_k, g_k) = L_{k+1}$. □

Lemma 5.13 implies that at most n iterations of the CG method are necessary to compute the solution to (5.13) because there are only n orthogonal vectors in $\mathbb{R}^n$.

Lemma 5.14. *Let us define $\delta_k = x^{k+1} - x^k$. If the CG method does not terminate at step k, then*

$$L_k = \mathrm{Lin}(\delta_0, \ldots, \delta_{k-1})$$

and

$$\delta_\ell^T H \delta_j = 0 \quad \text{for} \quad 0 \le \ell < j \le k-1.$$

Proof. Since $x^1 \in x^0 + L_1$, we have $x^1 = x^0 + \mu_0 g_0$ for some $\mu_0 \neq 0$. Hence $g_0 = (1/\mu)(x^1 - x^0)$. We obtain $L_1 = \mathrm{Lin}(x^1 - x^0)$.

Now suppose that the assertion is true for some k. We have $x^{k+1} \in x^0 + L_k$ and

$$\delta_k = x^{k+1} - x^k = x^{k+1} - x^0 + \sum_{i=0}^{k-1} \delta_i$$

Combined with $\delta_0, \ldots, \delta_{k-1} \in L_k \subset L_{k+1}$, we find that $\delta_k \in L_{k+1}$ and

$$\mathrm{Lin}(\delta_0, \ldots, \delta_k) \subset L_{k+1}. \tag{5.16}$$

Lemma 5.13 ensures that g_i is orthorgonal to L_i. Combined with $H\delta_k = H(x^{k+1} - x^k) = g_{k+1} - g_k$ and $\delta_j \in L_k$ for $j \le k-1$, we obtain

$$\delta_k^T H \delta_j = \delta_j^T (g_{k+1} - g_k) = 0.$$

Therefore, the vectors $\delta_0, \ldots, \delta_k$ are H-orthogonal and thus linearly independent. We conclude that equality must hold in (5.16). □

Using Lemmas 5.13 and 5.14, we are ready to provide an explicit iterative scheme for the CG method. Lemmas 5.13 and 5.14 ensure $L_{k+1} = \mathrm{Lin}(L_k, g_k)$ and $L_k = \mathrm{Lin}(\delta_0, \ldots, \delta_{k-1})$ with $\delta_k = x^{k+1} - x^k \in L_{k+1}$. Hence, we have

$$\delta_k = x^{k+1} - x^k = -\gamma_k g_k + \sum_{j=0}^{k-1} \lambda_j \delta_j$$

where $\lambda_j \in \mathbb{R}$, $\gamma_k \in \mathbb{R}^n$, and $g_j = \nabla f(x^j)$. Multiplying this identity by H and subsequently by δ_i with $0 \le i \le k-1$, we obtain

$$0 = \delta_k^T H \delta_i = -\gamma_k g_k^T H \delta_i + \sum_{j=0}^{k-1} \lambda_i \delta_i^T H \delta_j.$$

Using Lemma 5.14 and $H\delta_i = g_{i+1} - g_i$, we obtain

$$0 = -\gamma_k g_k^T H \delta_i + \lambda_i \delta_i^T H \delta_i = -\gamma_k g_k^T (g_{i+1} - g_i) + \lambda_i \delta_i^T H \delta_i.$$

Hence $\lambda_i = 0$ for $i < k-1$. For $i = k-1$, we have

$$\lambda_{k-1} = \frac{\gamma_k \|g_k\|_2^2}{\delta_{k-1}^T H \delta_{k-1}}.$$

This also ensures that $\gamma_k \neq 0$. We conclude that $x^{k+1} = x^k + \gamma_k d_k$, where

$$d_k = -g_k + (\lambda_{k-1}/\gamma_k)\delta_{k-1}.$$

We compute γ_k. Since $x^{k+1} = \operatorname{argmin}_{x \in x^0 + L_{k+1}} f(x)$, and $x^{k+1} = x^k + \gamma_k d_k$, we have

$$\gamma_k = \operatorname{argmin}_{\gamma \in \mathbb{R}} f(x^k + \gamma_k d_k).$$

We obtain (compare with Exercise 5.3),

$$\gamma_k = -\frac{d_k^T g_k}{d_k^T H d_k} = \frac{\|g_k\|_2^2}{d_k^T H d_k}.$$

Now we derive an identity for d_k that does not explicitly depend on δ_{k-1}. Using $\delta_{k-1} = \gamma_{k-1} d^{k-1}$ and the expression for γ_{k-1}, we obtain

$$\begin{aligned}
d_k &= -g_k + (\lambda_{k-1}/\gamma_k)\delta_{k-1} \\
&= -g_k + \frac{\|g_k\|_2^2}{\delta_{k-1}^T H \delta_{k-1}} \delta_{k-1}. \\
&= -g_k + (1/\gamma_{k-1}) \frac{\|g_k\|_2^2}{d_{k-1}^T H d_{k-1}} d_{k-1} \\
&= -g_k \frac{d_{k-1}^T H d_{k-1}}{\|g_{k-1}\|_2^2} \frac{\|g_k\|_2^2}{d_{k-1}^T H d_{k-1}} d_{k-1} \\
&= -g_k + \frac{\|g_k\|_2^2}{\|g_{k-1}\|_2^2} d_{k-1}.
\end{aligned}$$

We summarize our derivations in Algorithm 5.15.

Algorithm 5.15 (Conjugate gradient method)**.**

0. Choose initial point/starting point $x^0 \in \mathbb{R}^n$, compute $g_0 = \nabla f(x^0) = Hx^0 - b$, and define $d_0 = -g_0$.

For $k = 0, 1, 2, \ldots$

1. If $g_k = 0$, then terminate with x^k.
2. Compute $x^{k+1} = x^k + \gamma_k d^k$, where

$$\gamma_k = \frac{g_k^T g_k}{d_k^T H d_k}.$$

3. Compute $g_{k+1} = Hx^{k+1} - b$ and

$$d_{k+1} = -g_{k+1} + \beta_{k+1} d_k, \quad \text{where} \quad \beta_{k+1} = \frac{\|g_{k+1}\|_2^2}{\|g_k\|_2^2}.$$

We recall that the CG method requires at most n iterations until it terminates. We now state a result on the convergence rate of the CG method (see Algorithm 5.15).

Theorem 5.16. *If f is as in (5.14) and (x^k) are the iterates of Algorithm 5.15, then*

$$f(x^k) - f(x^*) \leq 4\left(\frac{\sqrt{\kappa_f} - 1}{\sqrt{\kappa_f} + 1}\right)^{2k} \big(f(x^0) - f(x^*)\big),$$

where κ_f is the condition number of f, that is, the ratio of the largest and the smallest eigenvalue of H.

Comparing this convergence rate with that of the gradient descent provided in Proposition 5.9, we have an improved dependence on the condition number κ_f. The convergence bound in (5.16) implies that if the condition number κ_f of the strongly convex quadratic function f is (very) small, then a fast convergence can be expected. However, this bound does not imply that a large condition number results in slow convergence of CG. The convergence of CG depends mostly on the eigenvalues of H; more precisely on the distribution of the eigenvalues of H. It can be shown that the CG method converges quickly if the eigenvalues of H are clustered.

Theorem 5.16 asserts that the CG method converges fast if the condition number of f is small.

5.4 Newton's method

We continue investigating methods the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n} f(x). \tag{5.17}$$

Throughout the section, we require $f\colon \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable.

In the simplest case, Newton's method intends to find a root of a differentiable function $\phi\colon \mathbb{R} \to \mathbb{R}$ with $\phi(t^*) = 0$ for some $t^* \in \mathbb{R}$. If t is close to t^*, then we have

$$0 = \phi(t^*) \approx \phi(t) + \phi'(t)(t^* - t).$$

The idea of Newton's method is to compute an approximation t_{k+1} to t^* via

$$0 = \phi(t_k) + \phi'(t_k)(t_{k+1} - t_k).$$

If $\phi'(t^k) \neq 0$, we can write

$$t_{k+1} = t_k - \phi'(t_k)^{-1}\phi(t_k).$$

Figure 5.5 provides an illustration.

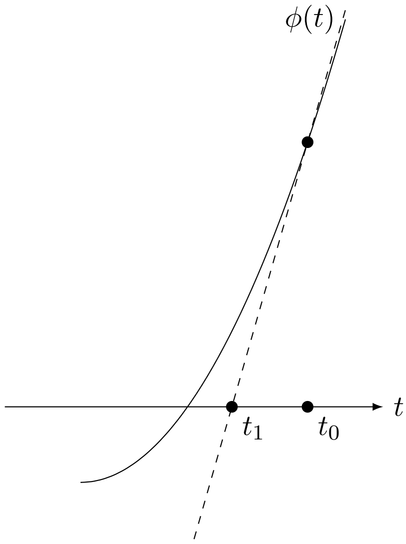

Figure 5.5: Illustration of Newton's method applied to the root finding of $\phi(t) = (1/2)t^2 - 1$. The figure shows the graph of ϕ (solid line) and the tangent (dashed line) of ϕ at t^0. The point t^1 is the zero of this tangent.

This iterative scheme also applies to foot finding of vector-valued mappings $F : \mathbb{R}^n \to \mathbb{R}^n$. If $F'(x^k)$ is an invertiable matrix, then the Newton iteration is given by

$$x^{k+1} = x^k - F'(x^k)^{-1}F(x^k).$$

For solving the unconstrained minimization problem (5.17), we apply Newton's method to the mapping $F = \nabla f$ and obtain the iterative scheme

$$x^{k+1} = x^k - \nabla^2 f(x^k)^{-1}\nabla f(x^k).$$

We may also motivate Newton's method via a second-order Taylor's expansion of f about x^k and subsequent minimization of the quadratic function. We have

$$f(x) \approx f(x^k) + \nabla f(x^k)(x - x^k) + (1/2)(x - x^k)^T\nabla^2 f(x^k)(x - x^k).$$

Let $\phi(x)$ be the quadratic function on the right-hand side. We obtain

$$\nabla\phi(x) = \nabla f(x^k) + \nabla^2 f(x^k)(x - x^k).$$

So computing x^{k+1} amounts to computing a stationary point of the model function ϕ.

Let us state the local Newton method as an algorithm.

Algorithm 5.17 (Local Newton method)**.**

0. Choose initial point/starting point $x^0 \in \mathbb{R}^n$.

For $k = 0, 1, 2, \ldots$

1. Compute s^k as a solution to
$$\nabla^2 f(x^k)s = -\nabla f(x^k).$$
2. Define $x^{k+1} = x^k + s^k$.

We recall that the second-order sufficient optimality conditions of minimizing f over $\mathbb{R}^n$ at a point x^* are given by

$$\nabla f(x^*) = 0 \text{ and there exists } \mu > 0 \text{ with } d^T\nabla^2 f(x^*)d \geq \mu d^T d \quad \text{for all} \quad d \in \mathbb{R}^n. \tag{5.18}$$

see Proposition 4.12. The latter is equivalent to $\nabla^2 f(x^*)$ is positive definite which is in turn equivalent to the fact that the minimum eigenvalue $\lambda_{\min}(\nabla^2 f(x^*))$ of $\nabla^2 f(x^*)$ is positive.

We state a basic convergence result of Algorithm 5.17.

Theorem 5.18. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be two times continuously differentiable in a neighborhood of $x^* \in \mathbb{R}^n$ and suppose that x^* satisfies the second-order sufficient optimality conditions* (5.18)*. Then for all x^0 sufficiently close to x^*, Algorithm* 5.17 *is well-defined and either computes $x^k = x^*$ for some k or generates a sequence (x^k) converging to x^* as $k \to \infty$ and the convergence rate is q-superlinear:*

$$\|x^{k+1} - x^*\|_2 = o(\|x^k - x^*\|_2).$$

We do not provide a proof of Theorem 5.18. However, we provide a short computation showing that if x^* fulfills the conditions in Theorem 5.18 and Algorithm 5.17 generates a sequence (x^k) converging to x^*, then we have q-superlinear convergence:

$$\|x^{k+1} - x^*\|_2 = o(\|x^k - x^*\|_2).$$

Let us verify the fast convergence. Since $\nabla^2 f$ is continuous and $x^k \to x^*$, we have $\nabla^2 f(x^k) \to \nabla^2 f(x^*)$. Combined with the fact that $\nabla^2 f(x^*)$ is positive definite, we find that $\nabla^2 f(x^k)$ is positive semidefinite for all sufficiently large k. Using Taylor's theorem, we find that

$$0 = \nabla f(x^*) = \nabla f(x^k) + \nabla^2 f(x^k)(x^* - x^k) + o(\|x^k - x^*\|_2).$$

Multiplying by the inverse of $\nabla^2 f(x^k)$ and using $\nabla^2 f(x^k)s^k = -\nabla f(x^k)$, we obtain

$$\begin{aligned} x^{k+1} - x^* = x^k - x^* + s^k = & \underbrace{x^k - x^*}_{=\nabla^2 f(x^k)^{-1}\nabla f(x^k) + o(\|x^k - x^*\|_2)} -\nabla^2 f(x^k)^{-1}\nabla f(x^k) \\ &= o(\|x^k - x^*\|_2). \end{aligned}$$

Hence, we obtain

$$\|x^{k+1} - x^*\|_2 = o(\|x^k - x^*\|_2).$$

We state and establish another convergence result of Algorithm 5.17. We require an error bound related to that in Lemma 5.5. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be twice differentiable. We say that $\nabla^2 f$ is Lipschitz continuous with Lipschitz constant $M \geq 0$ if for all x, $y \in \mathbb{R}^n$,

$$\|\nabla^2 f(y) - \nabla^2 f(x)\|_2 \leq M\|y - x\|_2.$$

The norm on the left-hand side is the spectral norm. For matrix $A \in \mathbb{R}^{n\times n}$, it is defined by

$$\|A\|_2 \coloneqq \sup_{\|x\|_2 \leq 1} \|Ax\|_2.$$

If $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite, then $\|A\|_2$ equals its maximum eigenvalue $\lambda_{\max}(A)$ and we have $\|A^{-1}\|_2 = 1/\lambda_{\min}(A)$. These facts may be established using an eigendecomposition of the symmetric positive definite matrix A. Moreover

$$\text{If} \quad A = A^T \in \mathbb{R}^{n\times n}, \quad \text{and} \quad B = B^T \in \mathbb{R}^{n\times n}, \quad \text{then} \quad |\lambda_{\min}(A) - \lambda_{\min}(B)| \leq \|A - B\|_2. \tag{5.19}$$

see, e.g., p. 408 in [9].

Lemma 5.19. *If $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and $\nabla^2 f$ is Lipschitz continuous with Lipschitz constant $M \geq 0$, then for all x, $y \in \mathbb{R}^n$,*

$$\|\nabla f(y) - \nabla f(x) - \nabla^2 f(x)(y - x)\|_2 \leq (M/2)\|y - x\|_2^2.$$

Proof. The proof is similar to that of Lemma 5.5 and hence omitted. □

Now we state and establish a convergence result of Algorithm 5.17. We say that a sequence (y^k) converges quadratically to $y^* \in \mathbb{R}^n$ if $y^k \to y^*$ and if there exists a constant $C > 0$ with

$$\|y^{k+1} - y^*\|_2 \leq C\|y^k - y^*\|_2^2. \tag{5.20}$$

Quadratic convergence is extremely fast. After a few iterations, each iteration approximately doubles the number of significant figures in y^{k+1}.

Theorem 5.20. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable and let $\nabla^2 f$ be Lipschitz continuous with Lipschitz constant $M > 0$. Suppose that x^* satisfies the second-order sufficient optimality conditions* (5.18). *If Algorithm* 5.17 *is initialized with an initial point $x^0 \in \mathbb{R}^n$ satisfying $\|x^0 - x^*\|_2 \leq 2\mu/(3M)$, then it generates a sequence (x^k) with $\|x^k - x^*\|_2 \leq 2\mu/(3M)$ and*

$$\|x^{k+1} - x^*\|_2 \leq 3M/(2\mu)\|x^k - x^*\|_2^2.$$

Proof. Let $y \in \mathbb{R}^n$ be a point with $\|y - x^*\|_2 \leq 2\mu/(3M)$. Let us first show that $\nabla^2 f(y)$ is positive definite. Since $\nabla^2 f$ is Lipschitz continuous with Lipschitz constant M, we have

$$\|\nabla^2 f(y) - \nabla^2 f(x^*)\|_2 \leq M\|y - x^*\|_2.$$

Hence (5.19) ensures

$$\lambda_{\min}(\nabla^2 f(y)) \geq \lambda_{\min}(\nabla^2 f(x^*)) - M\|y - x^*\|_2.$$

Combined with (5.18), we obtain

$$\lambda_{\min}(\nabla^2 f(y)) \geq \lambda_{\min}(\nabla^2 f(x^*)) - M\|y - x^*\|_2 \geq \mu - 2M\mu/(3M) = (1 - 2/3)\mu.$$

Hence the minimum eigenvalue of $\nabla^2 f(y)$ is $\geq (\mu/3)$ for all $y \in \mathbb{R}^n$ with $\|y - x^*\|_2 \leq 2\mu/(3M)$ and we have

$$\|\nabla^2 f(y)^{-1}\|_2 \leq (3/\mu).$$

Using Lemma 5.19 and $\nabla f(x^*) = 0$, we find that for all $y \in \mathbb{R}^n$,

$$\| - \nabla f(y) - \nabla^2 f(y)(x^* - y)\|_2 \leq (M/2)\|y - x^*|_2^2.$$

Using $x^{k+1} = x^k - \nabla^2 f(x^k)^{-1} \nabla f(x^k)$, we obtain

$$\begin{aligned} x^{k+1} - x^* &= x^k - x^* - \nabla^2 f(x^k)^{-1} \nabla f(x^k) \\ &= \nabla^2 f(x^k)^{-1} \big(\nabla^2 f(x^k)(x^k - x^*) - \nabla f(x^k)\big). \end{aligned}$$

Taking norms and applying our previous calculations, we obtain

$$\|x^{k+1} - x^*\|_2 \leq \|\nabla^2 f(x^k)^{-1}\|_2 \cdot (M/2)\|x^k - x^*\|_2^2 \leq 3M/(2\mu)\|x^k - x^*\|_2^2.$$

Since $\|x^0 - x^*\|_2 \leq 2\mu/(3M)$, we have $\|x^1 - x^*\|_2 \leq 3M/(2\mu) \cdot 2\mu/(3M)\|x^0 - x^*\|_2 \leq 2\mu/(3M)$. Inductively, we obtain $\|x^{k+1} - x^*\|_2 \leq \|x^k - x^*\|_2 \leq 2\mu/(3M)$. □

The local Newton method is affine invariant: Let $b \in \mathbb{R}^n$ and let $A \in \mathbb{R}^{n \times n}$ be an invertiable matrix. Let us apply the local Newton method to minimizing the function $g(y) := f(Ay + b)$. We have $\nabla g(y) = A^T \nabla f(Ay + b)$ and $\nabla^2 g(y) = A^T \nabla^2 f(Ay + b)A$. The Newton iteration applied to minimizing g reads

$$\begin{aligned} y^{k+1} &= y^k - \nabla^2 g(y^k)^{-1} \nabla g(y^k) \\ &= y^k - \left[A^T \nabla^2 f(Ay^k + b)A\right]^{-1} A^T \nabla f(Ay^k + b) \\ &= y^k - A^{-1} \nabla^2 f(Ay^k + b)^{-1} \left[(A^T)^{-1} A^T\right] \nabla f(Ay^k + b), \\ &= y^k - A^{-1} \nabla^2 f(Ay^k + b)^{-1} \nabla f(Ay^k + b), \end{aligned}$$

Multiplying by A and adding b, we obtain

$$Ay^{k+1} + b = Ay^k + b - \nabla^2 f(Ay^k + b)^{-1} \nabla f(Ay^k + b).$$

Defining $x^k = Ay^k + b$ and $x^{k+1} = Ay^{k+1} + b$, we find that this iteration is the same the Newton iteration applied to f.

The local Newton method for unconstrained minimization of f has several serious issues:

1. The matrices $\nabla^2 f(x^k)$ may not be invertible and therefore x^{k+1} cannot be computed in general.

2. The points x^{k+1} may not decrease the objective function f; it can happen that $f(x^{k+1}) > f(x^k)$.

3. The sequence (x^k) may diverge if x^0 is not close to x^*. Let us consider the local Newton method as applied to the one-dimensional optimization problem

$$\min_{t\in\mathbb{R}} \sqrt{1+t^2}.$$

   The optimal solution is $t^* = 0$. It can be shown that the Newton iteration is given by $t_{k+1} = -t_k^3$. If $|t_0| > 1$, the method diverges and if $|t_0| < 1$ it converges quickly. Moreover if $|t_0| = 1$, then Newton's method oscillates.

For these reasons, it is necessary to modify the local Newton method to make it a reliable minimization algorithm for unconstrained minimization.

Let us point out one modification of the local Newton method that is globally convergent for *strongly* convex objective functions f.

Algorithm 5.21 (A modified Newton method)**.**

0. Choose initial point/starting point $x^0 \in \mathbb{R}^n$.

For $k = 0, 1, 2, \dots$

1. Compute s^k as a solution to

$$\nabla^2 f(x^k)s = -\nabla f(x^k).$$

2. Compute γ_k via the minimization rule:

$$\gamma_k = \operatorname{argmin}_{\gamma\geq 0} f(x^k + \gamma_k s^k)$$

3. Define $x^{k+1} = x^k + \gamma_k s^k$.

If f is twice continuously differentiable, strongly convex, and the level set $\{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ is bounded, then the sequence (x^k) generated by Algorithm 5.21 converges to the unique minimizer of f. The modified Newton method (5.21) is well-defined for twice differentiable, strongly convex functions f, in this case $\nabla^2 f(x^k)$ is positive definite and hence invertiable. However, if f is not strongly convex, this algorithm may not be well-defined.

5.5 Variable metric methods*

We develop a somewhat general approach to developing minimization methods for unconstrained smooth minimization. Let us recall that the step $x^{k+1} = x^k - \gamma_k \nabla f(x^k)$ of the gradient method is the solution to

$$\min_{y \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T (y - x^k) + (1/(2\gamma_k))\|y - x^k\|_2^2,$$

provided that $\gamma_k > 0$. We can also view the update of the gradient method as $x^{k+1} = x^k + \gamma_k s^k$, where s^k solves

$$\min_{s \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T s + (1/2)\|s\|_2^2. \tag{5.21}$$

Let us verify that $x^{k+1} = x^k + \gamma_k s^k$ equals the update $x^{k+1} = x^k - \gamma \nabla f(x^k)$. Since s^k is the solution to (5.21), we have $\nabla f(x^k) + s^k = 0$. Hence $s^k = -\nabla f(x^k)$ and we obtain $x^{k+1} = x^k + \gamma_k s^k = x^k - \gamma_k \nabla f(x^k)$.

Let us assume that $\nabla^2 f(x^k)$ is positive definite. The iteration $x^{k+1} = x^k + s^k$ of the local Newton method is given by the solution s^k to the linear system $\nabla^2 f(x^k) s^k = -\nabla f(x^k)$. The Newton step s^k is also the solution to

$$\min_{s \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T s + (1/2) s^T \nabla^2 f(x^k) s.$$

One approach to motivating variable metric methods is to choose a symmetric positive definite matrix H_k and compute s^k as the solution to

$$\min_{s \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T s + (1/2) s^T H_k s \tag{5.22}$$

and use the update $x^{k+1} = x^k + \gamma_k s^k$, where the step size s^k is determined by line search. When choosing $H_k = I$, then we obtain the gradient method. If $\nabla^2 f(x^k)$ is positive definite, we can choose $H_k = \nabla^2 f(x^k)$ and $\gamma_k = 1$ to obtain the local Newton method.

The name "variable metric method" may be motivated as follows. If H_k is symmetric positive definite, then $(s, d)_{H_k} := s^T H_k d$ defines a scalar product and $\|x\|_{H_k} = \sqrt{x^T H_k x}$ a norm. The optimization problem (5.22) can be written as

$$\min_{s \in \mathbb{R}^n} f(x^k) + (H_k^{-1} \nabla f(x^k), s)_{H_k} + (1/2)\|s\|_{H_k}^2.$$

We state a generic variable metric method.

Algorithm 5.22 (Variable metric method)**.**

0. Choose initial point/starting point $x^0 \in \mathbb{R}^n$.

For $k = 0, 1, 2, \ldots$

1. Choose a symmetric positive definite matrix $H_k \in \mathbb{R}^{n \times n}$.

2. Compute s^k as a solution to
$$H_k s = -\nabla f(x^k). \tag{5.23}$$
3. Choose $\gamma_k \geq 0$ via line search applied to $\gamma \mapsto f(x^k + \gamma_k s^k)$.
4. Define $x^{k+1} = x^k + \gamma_k s^k$.

For the variable metric method to converge, it does not suffice to have the matrices H_k symmetric positive definite. We require them to satisfy the following condition: there exists $\mu > 0$ and $L > 0$ such that
$$\mu d^T d \leq d^T H_k d \leq L d^T d \quad \text{for all} \quad d \in \mathbb{R}^n, \quad k \in \mathbb{N}. \tag{5.24}$$
This is equivalent to requiring $\lambda_{\min}(H_k) \geq \mu$ and $\lambda_{\max}(H_k) \leq L$ for all $k \in \mathbb{N}$. This assumption requires the norms of H_k and H_k^{-1} to be bounded independently of the iteration counter k. In other words, the minimum eigenvalues of H_k must uniformly be bounded from below and the maximum eigenvalue of H_k must uniformly be bounded from above. This assumption prevents $\lambda_{\min}(H_k) \to 0$ and $\lambda_{\max}(H_k) \to \infty$, for example.

We state a global convergence result of the variable metric method.

Theorem 5.23. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and let the level set*
$$\{x \in \mathbb{R}^n \colon f(x) \leq f(x^0)\}$$
be bounded, where $x^0 \in \mathbb{R}^n$. We apply Algorithm 5.22 with starting value x^0 and suppose that either the minimization rule or the Armijo rule is used to compute the step sizes γ_k. Suppose that there exists constants $\mu > 0$ and $L \geq 0$ such that (5.24) holds true. Then the sequence (x^k) generated by Algorithm 5.22 is bounded, $f(x^{k+1}) \leq f(x^k)$ for all $k \in \mathbb{N}$, and all accumulation points of (x^k) are critical points of f.

The proof of this theorem may be established using arguments similar to those used in the proof of Theorem 5.3.

Let us discuss a few options for choosing the matrices H_k in Algorithm 5.22.

- The choice $H_k = I$ yields the gradient descent method.
- If f is strongly convex with parameter $\mu > 0$, then we can choose $H_k = \nabla^2 f(x^k)$. Since f is strongly convex with parameter μ, the first condition in (5.24) holds true. The upper bound holds true with L being the maximizer of $\|\nabla^2 f(x)\|_2$ over the level set $\{x \in \mathbb{R}^n \colon f(x) \leq f(x^0)\}$. The maximizer exists if f is twice continuously differentiable and the level set is bounded.
- The Levenberg–Marquardt regularization of the Hessian $\nabla^2 f(x^k)$ is the matrix
$$H_k = \nabla^2 f(x^k) + \rho_k I,$$
where $I \in \mathbb{R}^{n \times n}$ is the identity matrix and $\rho_k \geq 0$ is a regularization parameter. If ρ_k is larger than the minimum eigenvalue of $\nabla^2 f(x^k)$, then H_k is symmetric positive definite. Indeed, if $\rho_k > \lambda_{\min}(\nabla^2 f(x^k))$, then $\lambda_{\min}(H_k) = \lambda_{\min}(\nabla^2 f(x^k)) + \rho_k > 0$.

To ensure the conditions in (5.24), we can request that ρ_k is chosen in such a way that $\lambda_{\min}(H_k) \geq \mu$ for all $k \in \mathbb{N}$. Here $\mu > 0$ is some constant chosen before running the variable metric method.
The method of choice for checking whether a symmetric matrix is positive definite is the Cholesky decomposition. If applied to a symmetric matrix the decomposition succeeds, then this matrix is positive definite; if it fails, then the matrix is not positive definite. Using the Cholesky decomposition, we can compute rather efficiently, a small number $\rho_k \geq 0$ such that $H_k - \mu I$ is positive definite. This ensures $\lambda_{\min}(H_k) > \mu$. When the numbers ρ_k are properly chosen, then we can ensure also the second condition in (5.24) (provided that f is twice continuously differentiable and the level set of f corresponding to $f(x^0)$ is bounded.) However, this scheme may result in $H_k \neq \nabla^2 f(x^k)$ even if $\nabla^2 f(x^k)$ is positive definite.

The global convergence of the variable metric method can also be established under a milder condition on the symmetric positive definite matrices H_k than that in (5.24). We require instead of (5.24) the existence of a constant $C > 0$ such that

$$\kappa(H_k) = \frac{\lambda_{\max}(H_k)}{\lambda_{\min}(H_k)} \leq C \quad \text{for all} \quad k \in \mathbb{N}. \tag{5.25}$$

Here $\kappa(H_k)$ is the condition number of the symmetric positive definite matrix H_k. The condition (5.24) ensures (5.25) with $C = L/\mu$. However, (5.25) allows for a larger class of matrices H_k than (5.24). Let us define $s^k = -H_k^{-1}\nabla f(x^k)$. We have $-\nabla f(x^k)^T s^k = \nabla f(x^k)^T H_k^{-1} \nabla f(x^k)$. If $\nabla f(x^k) \neq 0$, then the condition (5.25) ensures

$$\frac{-\nabla f(x^k)^T s^k}{\|s^k\|_2} \geq (1/C)\|\nabla f(x^k)\|_2. \tag{5.26}$$

Let us verify this estimate.

$$-\nabla f(x^k)^T s^k \geq \nabla f(x^k)^T H_k^{-1} \nabla f(x^k) \geq \lambda_{\min}(H_k^{-1})\|\nabla f(x^k)\|_2^2 = (1/\lambda_{\max}(H_k))\|\nabla f(x^k)\|_2^2.$$

and

$$\|s^k\|_2 = \|H_k^{-1}\nabla f(x^k)\|_2 \leq \|H_k^{-1}\|_2 \|\nabla f(x^k)\|_2 = \lambda_{\min}(H_k)\|\nabla f(x^k)\|_2.$$

Combining these estimates and using (5.25), we obtain the assertion.

Before we state a convergence result of the variable metric method, let us introduce the Armijo rule (5.3) for steps s^k other than the gradient step $-\nabla f(x^k)$: Compute $\gamma_k > 0$ such that

$$\begin{aligned} f(x^k + \gamma_k s^k) &\leq f(x^k) + \varepsilon \gamma_k \nabla f(x^k)^T s^k, \\ f(x^k + \eta \gamma_k s^k) &\geq f(x^k) + \varepsilon \eta \gamma_k \nabla f(x^k)^T s^k, \end{aligned} \tag{5.27}$$

where $\varepsilon \in (0,1)$ and $\eta > 1$ are fixed parameters. For $s^k = -\nabla f(x^k)$, we obtain (5.3).

Theorem 5.24. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and let the level set*

$$S = \{x \in \mathbb{R}^n \colon f(x) \leq f(x^0)\}$$

be bounded, where $x^0 \in \mathbb{R}^n$. We apply Algorithm 5.22 with starting value x^0 and suppose the Armijo rule (5.27) is used to compute the step sizes γ_k. Suppose that there exists a constant $C > 0$ such that (5.25) holds true. Then the sequence (x^k) generated by Algorithm 5.22 is bounded, $f(x^{k+1}) \leq f(x^k)$ for all $k \in \mathbb{N}$, and all accumulation points of (x^k) are critical points of f.

Proof. Let us define $\phi(\gamma) \coloneqq f(x^k + \gamma s^k)$. We have $\phi'(0) = \nabla f(x^k)^T s^k$. Combined with $s^k = -H_k^{-1} \nabla f(x^k)$, we obtain $\phi'(0) = -\nabla f(x^k)^T H_k^{-1} \nabla f(x^k)$. So $\phi'(0) < 0$ provided that $\nabla f(x^k) \neq 0$. Using arguments similar to those in (5.1.1), we can show that there exists $\gamma_k > 0$ such that the Armijo conditions (5.27) hold true. We obtain $f(x^{k+1}) \leq f(x^k)$. We conclude that the sequence (x^k) is well-defined and contained in the level set S. Since $(f(x^k))$ is monotonically nonincreasing and $(x^k) \subset S$ is bounded, the sequence $(f(x^k))$ converges to some finite number as $k \to \infty$.

Let us assume that $\nabla f(x^k) \neq 0$ for all $k \in \mathbb{N}$. In light of (5.26), it suffices to show that

$$m_k \coloneqq \frac{-\nabla f(x^k)^T s^k}{\|s^k\|_2} \to 0 \quad \text{as} \quad k \to \infty.$$

Note that $m_k \geq 0$. Suppose that this term does not converge to 0. Then there exists $\delta > 0$ and a subsequence (m_{k_ℓ}) of (m_k) such that $m_{k_\ell} \geq \delta$ for all $\ell \in \mathbb{N}$. Using (5.27), we obtain for all ℓ,

$$f(x^k) - f(x^{k_\ell+1}) \geq -\varepsilon \gamma_{k_\ell} \frac{\nabla f(x^{k_\ell})^T s^{k_\ell}}{\|s^{k_\ell}\|_2} \|s^{k_\ell}\|_2 \geq \varepsilon \gamma_{k_\ell} \delta \|s^{k_\ell}\|_2.$$

Since $f(x^k) - f(x^{k+1}) \to 0$ as $k \to \infty$, we have $\gamma_{k_\ell} \|s^{k_\ell}\|_2 \to 0$ as $\ell \to \infty$. Hence there exists $r_1 \in (0, \infty)$ such that $\gamma_{k_\ell} \|s^{k_\ell}\|_2 \leq r_1$ for all $\ell \in \mathbb{N}$. Since the level set S is bounded and closed (as f is continuous), the set $X \coloneqq \{x + s \colon x \in S, \|s\|_2 \leq \eta r_1\}$ is bounded and closed. Consequently, $\mathrm{Conv}(X)$ is bounded, closed, and convex (see Exercise 1.6). The function $x \mapsto \|\nabla^2 f(x)\|_2$ is continuous and hence there exists $L > 0$ such that $\|\nabla^2 f(x)\|_2 \leq L$ for all $x \in \mathrm{Conv}(X)$. Consequently, ∇f is Lipschitz continuous on $\mathrm{Conv}(X)$ with Lipschitz constant L. Using a generalization of Lemma 5.5, we obtain for all x, $y \in \mathrm{Conv}(X)$,

$$f(y) - f(x) - \nabla f(x)^T (y - x) \leq (L/2) \|y - x\|_2^2$$

Using (5.27), we further obtain

$$f(x^{k_\ell} + \eta \gamma_{k_\ell} s^{k_\ell}) - f(x^{k_\ell}) \geq \varepsilon \eta \gamma_{k_\ell} \nabla f(x^{k_\ell})^T s^{k_\ell}.$$

We bound the left-hand side using the above inequality and $m_{k_\ell} \geq \delta$. We get

$$f(x^{k_\ell} + \eta \gamma_{k_\ell} s^{k_\ell}) - f(x^{k_\ell}) \leq \eta \gamma_{k_\ell} \nabla f(x^{k_\ell})^T s^{k_\ell} + (L/2) \gamma_{k_\ell}^2 \eta^2 \|s^{k_\ell}\|_2^2.$$

Combining these two inequalities,

$$
\begin{aligned}
(L/2)\eta^2\gamma_{k_\ell}^2\|s^{k_\ell}\|_2^2. &\geq (\varepsilon-1)\eta\gamma_{k_\ell}\nabla f(x^{k_\ell})^T s^{k_\ell} = -(1-\varepsilon)\eta\gamma_{k_\ell}\frac{\nabla f(x^{k_\ell})^T s^{k_\ell}}{\|s^{k_\ell}\|_2}\|s^{k_\ell}\|_2 \\
&\geq (1-\varepsilon)\eta\gamma_{k_\ell}\delta\|s^{k_\ell}\|_2.
\end{aligned}
$$

Dividing by $\eta\gamma_{k_\ell}\|s^{k_\ell}\|_2$, we have $(L/2)\eta\gamma_{k_\ell}\|s^{k_\ell}\|_2 \geq (1-\varepsilon)\delta$. Taking limits as $\ell \to 0$, we obtain $0 \geq (1-\varepsilon)\delta$. Since $\varepsilon \in (0,1)$, we obtain a contraction and hence $m_k \to 0$ as $k \to \infty$.

Now, let $\bar{x}$ be an accumulation point of (x^k) and let (x^{k_ℓ}) be a subsequence of (x^k) converging to $\bar{x}$ as $\ell \to \infty$. If $\nabla f(x^\ell) = 0$ for some ℓ, then we have $x^{k_\ell} = x^{k_\ell+1} = \cdots = x^{k_{\ell+1}} = \cdots$ and hence $x^\ell = \bar{x}$. Now let $\nabla f(x^{k_\ell}) \neq 0$ for all all $\ell \in \mathbb{N}$. Using $m_k \to 0$ and (5.26), we find that $\nabla f(x^k) \to 0$. Since ∇f is continuous, we must have $\nabla f(\bar{x}) = 0$. □

5.6 Quasi-Newton methods

Quasi-Newton methods can be interpreted as variable metric methods (see Algorithm 5.22). Instead of using the Hessian matrix $\nabla^2 f(x^{k+1})$, quasi-Newton methods construct certain approximations H_{k+1} to it. These approximations are constructed using certain update formulas and are required to satisfy the quasi-Newton equation

$$
H_{k+1}(x^{k+1} - x^k) = \nabla f(x^{k+1}) - \nabla f(x^k). \tag{5.28}
$$

The Hessian matrix $\nabla^2 f(x^{k+1})$ may not satisfy this relation. However, if f is a strongly convex quadratic function, say, $f(x) = (1/2)x^T Ax - b^T x$, then $H_{k+1} = \nabla^2 f(x^{k+1})$ satisfies the quasi-Newton equation. The quasi-Newton equation may be motivated by considering a first-order Taylor's expansion of the gradient ∇f at x^k about the expansion point x^{k+1}. We have

$$
\nabla f(x^{k+1}) - \nabla f(x^k) \approx \nabla^2 f(x^k)(x^{k+1} - x^k).
$$

We can always find a matrix H_{k+1} that satisfies the quasi-Newton equation:

$$
\nabla f(x^{k+1}) - \nabla f(x^k) = G_{k+1}(x^{k+1} - x^k),
$$

where

$$
G_{k+1} = \int_0^1 \nabla^2 f(x^k + \tau(x^{k+1} - x^k))d\tau.
$$

However, this matrix is generally not available computationally, as it is defined via an integral.

As already mentioned, quasi-Newton matrices are defined using update formulas. These update formulas use H_k and the differences $x^{k+1} - x^k$, and $\nabla f(x^{k+1}) - \nabla f(x^k)$ to obtain H_{k+1}. For $k = 0$, we get to choose x^0 and a symmetric positive definite matrix H_0.

We may require the following conditions on potential update formulas:

1. H_{k+1} is symmetric and positive definite,
2. H_{k+1} satisfies the quasi-Newton equation, and approximates $\nabla^2 f(x^{k+1})$, and
3. the computational costs for updating and solving linear systems involving H_{k+1} are low.

Let us attempt to construct an update rule that satisfies our requirements. Let us define

$$d^k := x^{k+1} - x^k, \quad \text{and} \quad p^k := \nabla f(x^{k+1}) - \nabla f(x^k).$$

We consider the update

$$H_{k+1} = H_k + \rho_k u^k (u^k)^T, \tag{5.29}$$

where $\rho_k \in \mathbb{R}$ and $u^k \in \mathbb{R}^n$ is a vector with $\|u^k\|_2 = 1$. The matrix $u^k(u^k)^T$ has rank one and is called the outer product of the vector u^k. (The inner product is $(u^k)^T u^k$.) So H_{k+1} is a rank one update of H_k. This is the simplest possible update of a matrix. We require that H_{k+1} satisfies the quasi-Newton equation (5.28) and based on this requirement we compute ρ_k and u^k. Inserting H_{k+1} into the quasi-Newton equation, we obtain

$$H_{k+1} d^k = H_k d^k + \rho_k \big(u^k (u^k)^T\big) d^k = p^k. \tag{5.30}$$

Our task is to provide formulas for ρ_k and u^k. If $p^k - H_k d^k = 0$, and $\nabla f(x^k) + H_k d^k = 0$ (that is, d^k solves linear system in (5.23)), then we obtain

$$\nabla f(x^{k+1}) = \nabla f(x^k) + p^k = \nabla f(x^k) + H_k d^k = 0.$$

This means x^{k+1} is a stationary point of f. If $p^k - H_k d^k \neq 0$, then we obtain from (5.30) the identity

$$H_k d^k - p^k = -\rho_k (u^k (u^k)^T) d^k = -\rho_k ((d^k)^T u^k) u^k. \tag{5.31}$$

The latter identity can be established using computations. Since $-\rho_k((d^k)^T u^k)$ is a real number and the right-hand side is nonzero, we find that the vector u^k is a multiple of the vector $H_k d^k - p^k$:

$$u^k = \pm \frac{H_k d^k - p^k}{\|H_k d^k - p^k\|_2}.$$

Let us choose

$$u^k = \frac{H_k d^k - p^k}{\|H_k d^k - p^k\|_2}.$$

Now we compute ρ_k. If $(d^k)^T (H_k d^k - p^k) \neq 0$, then (5.31) ensures

$$\|H_k d^k - p^k\|_2 (d^k)^T u_k = -\rho_k ((d^k)^T u^k) u^k.$$

Hence

$$\|H_k d^k - p^k\|_2 = -\rho_k (d^k)^T u^k = -\rho_k \frac{(d^k)^T (H_k d^k - p^k)}{\|H_k d^k - p^k\|_2},$$

yielding

$$\rho_k = \frac{\|H_k d^k - p^k\|_2^2}{(d^k)^T (p^k - H_k d^k)}.$$

Putting together the pieces and using (5.29), we obtain

$$H_{k+1} = H_k + \frac{(p^k - H_k d^k)(p^k - H_k d^k)^T}{(p^k - H_k d^k)^T d^k}.$$

These update formula is called *symmetric rank-one (SR1)* formula. It has many issues: (i) we can have $(d^k)^T (H_k d^k - p^k) = 0$ and then the update formula is not well-defined; (ii) if $(d^k)^T (H_k d^k - p^k) < 0$, then H_{k+1} may not be positive definite even if H_k is positive definite; (iii) H_{k+1} may lack an inverse; and (iv) the vector $-H_{k+1}^{-1} \nabla f(x^{k+1})$ may not be a descent direction of f.

For these reasons the SR1 update formula violates some of our requirements. The SR update formula is a rank-one update formula of H_k. The most successful update formulas are given by rank-two updates of H_k:

$$H_{k+1} = H_k + \rho_k u^k (u^k)^T + \varrho_k v^k (v^k)^T.$$

with ρ_k, $\varrho_k \in \mathbb{R}$ and u^k and v^k are nonzero vectors. We state four rank-two update formulas:

- Broyden–Fletcher–Goldfarb–Shanno (BFGS) update formula:

$$H_{k+1}^{\mathrm{BFGS}} = H_k + \frac{p^k (p^k)^T}{(p^k)^T d^k} - \frac{H_k d^k (H_k d^k)^T}{(d^k)^T H_k d^k}.$$

- Davidon–Fletcher–Powell (DFP) update formula:

$$H_{k+1}^{\mathrm{DFP}} = H_k + \frac{(p^k - H_k d^k)(p^k)^T + p^k (p - H_k d^k)^T}{(p^k)^T d^k} - \frac{(p^k - H_k d^k)^T d^k}{((p^k)^T d^k)^2} p^k (p^k)^T$$

- Broyden family:

$$H_{k+1}^{\lambda} = (1-\lambda) H_{k+1}^{\mathrm{BFGS}} + \lambda H_{k+1}^{\mathrm{DFP}} \quad \text{with} \quad \lambda \in \mathbb{R}.$$

- convex Broyden family: H_{k+1}^{λ} with $\lambda \in [0,1]$.

  For $\lambda = 0$, we obtain the BFGS update formula.

Let us discuss basic properties of the update formulas given by the Broyden family.

Proposition 5.25. 1. *If H_k is symmetric, $(p^k)^T d^k \neq 0$ and $(d^k)^T H_k d^k \neq 0$, then the matrices H_{k+1}^{λ}, $\lambda \in \mathbb{R}$, are well-defined, symmetric and satisfy the quasi-Newton equation* (5.28).

2. *If H_k is symmetric positive definite and $(p^k)^T d^k > 0$, then the matrices H_{k+1}^{λ}, $\lambda \geq 0$, are symmetric positive definite.*

The proposition shows that we can use the BFGS update formula within a variable metric method, provided that H_0 is symmetric positive definite and $(p^k)^T d^k > 0$ for all iterations $k \in \mathbb{N}$. However, this approach would require us to solve linear systems involving the BFGS matrices. Instead of constructing the matrices H_{k+1}^{BFGS}, it would be helpful to construct their inverses B_{k+1}^{BFGS}, as this would only require matrix-vector multiplications for step computations. Before we provide update rules for inverse quasi-Newton matrices, let us briefly show how we can use the minimization rule to ensure $(p^k)^T d^k > 0$. Let H_k be symmetric positive definite, let f be twice continuously differentiable and bounded from below, and $\nabla f(x^k) \neq 0$. We define $s^k = -H_k^{-1}\nabla f(x^k)$ and compute γ_k via minimization rule applied to $\gamma \mapsto \phi(\gamma) = f(x^k + \gamma s^k)$. Since H_k is symmetric positive definite and $\nabla f(x^k) \neq 0$, s^k is a descent direction of f. The minimization rule computes $\gamma_k > 0$. We have $\phi'(\gamma_k) = 0$ and $\phi'(\gamma_k) = \nabla f(x^k + \gamma_k s^k)^T s^k$. Since $d^k = x^{k+1} - x^k = \gamma_k s^k$, we obtain

$$(p^k)^T d^k = \gamma_k (p^k)^T s^k = \underbrace{\gamma_k \nabla f(x^{k+1})^T s^k}_{=0} - \gamma_k \nabla f(x^k)^T s^k = \gamma_k \nabla f(x^k)^T H_k^{-1} \nabla f(x^k) > 0.$$

Let us provide the inverse update formulas for the BFGS and DFP update formulas.

Proposition 5.26. *Let H_k be symmetric positive definite and define $B_k = H_k^{-1}$. Suppose that $(p^k)^T d^k > 0$. Then the inverse updates for the matrices $B_{k+1}^{BFGS} = (H_{k+1}^{BFGS})^{-1}$ and $B_{k+1}^{DFP} = (H_{k+1}^{DFP})^{-1}$ are given by*

$$\begin{aligned} B_{k+1}^{BFGS} &= B_k + \frac{(d^k - B_k p^k)(d^k)^T + d^k(d^k - B_k d^k)^T}{(d^k)^T p^k} - \frac{(d^k - B_k p^k)^T p^k}{((d^k)^T p^k)^2} d^k (d^k)^T, \quad (5.32)\\ B_{k+1}^{DFP} &= B_k + \frac{d^k (d^k)^T}{(d^k)^T p^k} - \frac{B_k p^k (B_k p^k)^T}{(p^k)^T B_k p^k}. \end{aligned}$$

We now state a quasi-Newton method, the BFGS method. It uses the inverse BFGS updates in (5.32).

Algorithm 5.27 (Inverse BFGS method)**.**

0. Choose initial point/starting point $x^0 \in \mathbb{R}^n$ and a symmetric positive definite matrix B_0.

For $k = 0, 1, 2, \ldots$

1. Compute
$$s^k = -B_k \nabla f(x^k)$$
2. Choose $\gamma_k \geq 0$ via line search applied to $\gamma \mapsto f(x^k + \gamma s^k)$.
3. Define $x^{k+1} = x^k + \gamma_k s^k$.
4. Define $d^k = x^{k+1} - x^k$ and $p^k = \nabla f(x^{k+1}) - \nabla f(x^k)$
5. Compute B_{k+1} using (5.32).

We close the section with a few remarks about quasi-Newton methods and Algorithm 5.27.

- In Algorithm 5.27, we have not specified a line search procedure. In practical implementations, the step size γ_k is implemented via the Wolfe line search: Choose $\gamma_k > 0$ such that
$$\begin{aligned} f(x^k + \gamma_k s^k) &\leq f(x^k) + c_1 \gamma_k \nabla f(x^k)^T s^k, \\ \nabla f(x^k + \gamma_k s^k)^T s^k &\geq c_2 \nabla f(x^k)^T s^k, \end{aligned}$$
where $0 < c_1 < c_2 < 1$. The first condition ensures decrease in f, provided that $\nabla f(x^k)^T s^k < 0$ and the second condition ensures that γ_k is "large enough."

  The Wolfe line search can be used to show that $(d^k)^T p^k > 0$, a requirement for the matrices B_{k+1} be positive definite.
- Let f be twice continuously differentiable, strongly convex and ∇f be Lipschitz continuous. If Algorithm 5.27 is used with the Wolfe line search and $\{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ is bounded, then it can be shown that the sequence generated by Algorithm 5.27 converges to the unique minimizer of f.
- Global convergence to stationary points of quasi-Newton methods and in particular Algorithm 5.27 may not occur for general nonconvex objective functions. Counterexamples are known. In light of Theorem 5.24, the issue appears to be that the condition number of the matrices B_k may not be uniformly bounded.

  However, under certain assumptions, it is possible to show that the sequence (x^k) generated by Algorithm 5.27 converges superlinearly to x^*, provided that (x^k) converges to x^* and x^* is a stationary point of f with positive definite Hessian at x^*.
- Very recently, convergence rates for quasi-Newton method applied to strongly convex quadratic functions have been established.
- If Algorithm 5.27 is used with the inverse DFP update formulas, and applied to strongly convex quadratic functions, then it computes the solution in no more than n steps. If $B_0 = I$, then this scheme is the same as the conjugate gradient method.

- The efficient implementation of quasi-Newton method requires many considerations to obtain reliable implementations.
- The update formula (5.32) can also be written as

$$B_{k+1}^{\mathrm{BFGS}} = V_k^T B_k V_k + \rho_k d^k (d^k)^T,$$

where

$$\rho_k = \frac{1}{(p^k)^T d^k}, \quad V_k = I - \rho_k p^k (d^k)^T.$$

The matrices $B_{k+1} = B_{k+1}^{\mathrm{BFGS}}$ are generally dense and hence the cost of storing them can be large if n is large. For this reason, limited memory BFGS (L-BFGS) methods have been developed. The update of H_0 within L-BFGS is performed using p^i and d^i for $k - \min\{k, m-1\} \le i \le k$ for some small natural number m instead of p^i and d^i for $0 \le i \le k$.

5.7 Cubic Regularization of Newton's method

The cubic Newton method has been a breakthrough in nonlinear optimization. To motivate the method, let us recall that the Newton step s^k is a critical point of the optimization problem

$$\min_{s \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T s + (1/2) s^T \nabla^2 f(x^k) s.$$

The quadratic objective function provides an approximation to $f(x^k + s)$:

$$f(x^k + s) \approx f(x^k) + \nabla f(x^k)^T s + (1/2) s^T \nabla^2 f(x^k) s.$$

We can quantify the error of the second order Taylor's expansion using the next lemma.

Lemma 5.28. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be twice differentiable. If $\nabla^2 f$ is Lipschitz continuous with Lipschitz constant $L \ge 0$, then for all x, $y \in \mathbb{R}^n$,*

$$|f(y) - f(x) - \nabla f(x)^T (y-x) + (1/2)(y-x)^T \nabla^2 f(x)(y-x)| \le (L/6)\|y-x\|_2^3.$$

Proof. The verification is related to that of Lemma 5.5. □

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be twice differentiable and let $\nabla^2 f$ be Lipschitz continuous with Lipschitz constant $L > 0$. In its simplest form, the cubic Newton method computes $M_k > 0$ and s^k as a minimizer of the cubic subproblem

$$\min_{s \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T s + (1/2) s^T \nabla^2 f(x^k) s + (M_k/6)\|s\|_2^3, \tag{5.33}$$

and subsequently defines $x^{k+1} = x^k + s^k$. If $M_k \ge L$, then the objective function

$$\phi_k(s) \coloneqq f(x^k) + \nabla f(x^k)^T s + (1/2) s^T \nabla^2 f(x^k) s + (M_k/6)\|s\|_2^3.$$

of this auxiliary problem provides a global upper bound on $f(x^k + s)$ according to Lemma 5.28, that is, $f(x^k + s) \le \phi_k(s)$ for all $s \in \mathbb{R}^n$. If $M_k \ge L$ and $s^k \ne 0$, then we have $f(x^k + s^k) \le \phi_k(s^k) < \phi_k(0) = f(x^k)$. If $s^k = 0$, then we can show that $\nabla f(x^k) = 0$ and that $\nabla^2 f(x^k)$ is positive semidefinite (see Proposition 5.30).

We formulate the cubic Newton method.

Algorithm 5.29 (Cubic Newton method)**.**

0. Choose initial point/starting point $x^0 \in \mathbb{R}^n$ and $L_0 > 0$ with $L_0 \le 2L$.

For $k = 0, 1, 2, \ldots$

1. Compute $M_k \in [L_0, 2L]$ such that

$$f(x^k + s^k) \le \phi_k(s^k), \quad \text{where} \quad s^k \quad \text{solves} \quad (5.33).$$

2. Define $x^{k+1} = x^k + s^k$.

The first step of the cubic Newton method may be implemented as follows. Fix an iteration $k \in \mathbb{N}$ and choose $M_k \ge L_0$. We compute the solution s^k to (5.33). If M_k does not fulfill $f(x^k + s^k) \le \phi_k(s^k)$, we increase M_k by two. We define $M_k = 2M_k$ and repeat this scheme until $f(x^k + s^k) \le \phi_k(s^k)$. Since $\nabla^2 f$ is Lipschitz continuous with Lipschitz constant $L > 0$, we have $M_k \le 2L$. Once we have found M_k such that $f(x^k + s^k) \le \phi_k(s^k)$, we define $M_{k+1} \coloneqq \max\{L_0, M_k/2\}$ as our estimate of L for the next iteration. This definition ensures that our estimate of L may also be decreased. The outlined approach is one method to implement the first step of the cubic Newton method. Other schemes are possible.

Our next goal is to establish convergence rates for the cubic Newton method. We now analyze basic properties of the cubic subproblem.

Proposition 5.30. *Let $M > 0$, let $g \in \mathbb{R}^n$, and let $H \in \mathbb{R}^{n\times n}$ be a symmetric matrix. We consider*

$$\min_{s\in\mathbb{R}^n} g^T s + (1/2)s^T H s + (M/6)\|s\|_2^3. \tag{5.34}$$

Then

1. *The problem* (5.34) *has a global solution.*
2. *If s^* is a global solution to* (5.34)*, then*

$$\begin{aligned} &g + Hs^* + (M/2)\|s^*\|_2 s^* = 0, \quad \textit{and} \\ &H + (M/2)\|s^*\|_2 I \quad \textit{is positive semidefinite.} \end{aligned} \tag{5.35}$$

3. *If s^* satisfies* (5.35)*, then*

$$g^T s^* + (1/2)(s^*)^T H s^* + (M/6)\|s^*\|_2^3 \le -(M/12)\|s^*\|_2^3. \tag{5.36}$$

Proof. Let us define $\phi(s) = g^T s + (1/2)s^T Hs + (M/6)\|s\|_2^3$. This function is twice continuously differentiable and we have

$$\nabla\phi(s) = Hs + (M/2)\|s\|_2 s, \quad \nabla^2\phi(s) = H + (M/2)\|s\|_2 I + \frac{M}{2\|s\|_2} ss^T.$$

1. Since $M > 0$, we have $\phi(s) \to \infty$ as $\|s\|_2 \to \infty$. Hence the level set $\{s \in \mathbb{R}^n : \phi(s) \leq \phi(s^0)\}$ for all $s^0 \in \mathbb{R}^n$ is bounded. It is also closed as ϕ is continuous. Consequently, ϕ has a global minimizer.

2. Let s^* be a global solution to (5.34). Then the second-order necessary optimality conditions for unconstrained minimization ensure $\nabla\phi(s^*) = 0$ and that $\nabla^2\phi(s^*)$ is semi-positive definite. Now let $d \in \mathbb{R}^n$. We have to show that $d^T(H + (M/2)\|s^*\|_2 I)d \geq 0$. If $d^T s^* = 0$, then we have $0 \leq d^T\nabla^2\phi(s^*)d = d^T(H + (M/2)\|s^*\|_2 I)d$. Now let $d^T s^* < 0$ and define

$$t = -\frac{2d^T s^*}{\|d\|_2^2}.$$

We have $t > 0$ and

$$\|s^* + td\|_2^2 = \|s^*\|_2^2 + 2td^T s^* + t^2\|d\|_2^2 = \|s^*\|_2^2.$$

Let us also define $q(s) = g^T s + (1/2)s^T Hs$. We have $\phi(s) = q(s) + (M/6)\|s\|_2^3$. Since s^* is a global solution to (5.34), and $\|s^* + td\|_2 = \|s^*\|_2$, we have

$$0 \leq \phi(s^* + td) - \phi(s^*) = q(s^* + td) - q(s^*).$$

Since $\nabla\phi(s^*) = 0$, we also have

$$(g + Hs^*)^T d = -(M/2)\|s^*\|_2 d^T s^*.$$

A second-order Taylor's expansion of q about s^* yields

$$\begin{aligned}
0 \leq q(s^* + td) - q(s^*) &= t\nabla q(s^*)^T d + (1/2)t^2 d^T Hd \\
&= t(g + Hs^*)^T d + (1/2)t^2 d^T Hd \\
&= -t(M/2)\|s^*\|_2 d^T s^* + (1/2)t^2 d^T Hd \\
&= t^2(M/4)\|s^*\|_2\|d\|_2^2 + (1/2)t^2 d^T Hd \\
&= (1/2)t^2 d^T(H + (M/2)\|s^*\|_2 I)d.
\end{aligned}$$

Since $t > 0$, we deduce $d^T(H + (M/2)\|s^*\|_2 I)d \geq 0$ for all $d^T s^* \leq 0$. Since $(-d)^T(H + (M/2)\|s^*\|_2 I)(-d) = d^T(H + (M/2)\|s^*\|_2 I)d$, we obtain $d^T(H + (M/2)\|s^*\|_2 I)d \geq 0$ for all $d \in \mathbb{R}^n$. Putting together the pieces, we obtain (5.35).

3. Using (5.35), we compute

$$\begin{aligned}
\phi(s^*) &= (-Hs^* - (M/2)\|s^*\|_2 s^* I)^T s^* + (1/2)(s^*)^T Hs^* + (M/6)\|s^*\|_2^3 \\
&= -(1/2)(s^*)^T(H + (M/2)\|s^*\|_2 s^* I)^T s^* - (M/12)\|s^*\|_2^3 \\
&\leq -(M/12)\|s^*\|_2^3.
\end{aligned}$$

□

It can be shown that if s^* satisfies (5.35), then it is a global solution to (5.34). Therefore, the conditions in (5.35) are necessary and sufficient optimality conditions.

We are now ready to establish convergence rates of the cubic Newton method. Let us define the error measure

$$\mu(x) := \max\left\{\sqrt{\frac{2}{3L}\|\nabla f(x)\|_2}, -\frac{1}{2L}\lambda_{\min}(\nabla^2 f(x))\right\}.$$

The function μ measures in some sense the violation of the second-order necessary optimality conditions for unconstrained minimization. In particular, x^* satisfies the second-order necessary optimality conditions $\nabla f(x^*) = 0$ and $\lambda_{\min}(\nabla^2 f(x^*)) \geq 0$ if and only if $\mu(x^*) = 0$. We also have for all $x \in \mathbb{R}^n$,

$$\|\nabla f(x)\|_2 \leq \frac{3L}{2}\mu(x)^2 \quad \text{and} \quad \lambda_{\min}(\nabla^2 f(x)) \geq -2L\mu(x).$$

Theorem 5.31. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable and let $\nabla^2 f$ be Lipschitz continuous with Lipschitz constant $L > 0$. Suppose that f is bounded from below by f^*, that is, $f(x) \geq f^*$ for all $x \in \mathbb{R}^n$. Then Algorithm* 5.29 *generates a sequence (x^k) with $f(x^{k+1}) \leq f(x^k)$ for all $k \in \mathbb{N}$ and for all $K \in \mathbb{N}$,*

$$\min_{0\leq k\leq K} \mu(x^k) \leq \left(\frac{12}{L_0}\frac{f(x^0) - f^*}{K}\right)^{1/3}.$$

Proof. Using (5.36), we find that

$$f(x^{k+1}) = f(x^k + s^k) \leq \phi_k(s^k) \leq f(x^k) - (M_k/12)\|s^k\|_2^3 \leq f(x^k).$$

Combined with $M_k \geq L_0$,

$$f(x^0) - f^* \geq \sum_{k=0}^{K}[f(x^k) - f(x^{k+1})] \geq \sum_{k=0}^{K}\frac{M_k}{12}\|s^k\|_2^3 \geq \frac{L_0}{12}\cdot(K+1)\min_{0\leq k\leq K}\|s^k\|_2^3.$$

Using (5.35), we have

$$\nabla f(x^{k+1}) = \nabla f(x^{k+1}) \underbrace{-\nabla f(x^k) - \nabla^2 f(x^k)s^k - (M_k/2)\|s^k\|_2 s^k}_{=0 \text{ (see (5.35))}}.$$

Combined with Lemma 5.19, $x^{k+1} = x^k + s^k$, $M_k \leq 2L$, and the triangle inequality, we find that

$$\|\nabla f(x^{k+1})\|_2 \leq (L/2)\|s^k\|_2^2 + (M_k/2)\|s^k\|_2^2 \leq (L/2)\|s^k\|_2^2 + L\|s^k\|_2^2 = (3/2)L\|s^k\|_2^2.$$

Hence

$$(2/(3L))\|\nabla f(x^{k+1})\|_2 \leq \|s^k\|_2^2.$$

The optimality conditions in (5.35) further ensure that

$$\nabla^2 f(x^k) + (M_k/2)\|s^k\|_2 I$$

is positive definite. Moreover, $\|\nabla^2 f(x^{k+1}) - \nabla^2 f(x^k)\|_2 \leq L\|s^k\|_2$. Combined with (5.19), and $M_k \leq 2L$,

$$\lambda_{\min}(\nabla^2 f(x^{k+1})) \geq \lambda_{\min}(\nabla^2 f(x^k)) - L\|s^k\|_2 \geq -L\|s^k\|_2 - (M_k/2)\|s^k\|_2 \geq -2L\|s^k\|_2.$$

In other words,

$$-(1/(2L))\lambda_{\min}(\nabla^2 f(x^{k+1})) \leq \|s^k\|_2.$$

Hence

$$\mu(x^{k+1}) = \max\left\{\sqrt{\frac{2}{3L}\|\nabla f(x^{k+1})\|_2}, -\frac{1}{2L}\lambda_{\min}(\nabla^2 f(x^{k+1}))\right\} \leq \|s^k\|_2.$$

Combined with

$$\left(\frac{12}{L_0}\frac{f(x^0) - f^*}{K+1}\right)^{1/3} \geq \min_{0\leq k\leq K} \|s^k\|_2.$$

We obtain

$$\min_{0\leq k\leq K} \mu(x^{k+1}) \leq \left(\frac{12}{L_0}\frac{f(x^0) - f^*}{K+1}\right)^{1/3}.$$

□

Theorem 5.31 ensures in particular that $\min_{0\leq i\leq k} \|\nabla f(x^i)\|$ decreases with a rate proportional to $1/k^{2/3}$. The gradient method has the much slower convergence rate $1/k^{1/2}$ (see Theorem 5.6). Using Theorem 5.31, we can show that each accumulation point x^* of a sequence generated by the cubic Newton method (x^k) satisfies the second-order necessary optimality conditions: $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite. Theorem 5.31 can be generalized to requiring $\nabla^2 f$ being Lipschitz continuous only on the level set $\{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$.

The cubic Newton method requires the solution of the subproblems (5.33). It turns out that these subproblems can efficiently be solved to global optimality. This is a deep fact. Figure 5.6 depicts a graph of an objective function of the cubic subproblem. This function has one local, nonglobal minimizer. Generally, the cubic subproblem can have many local, nonglobal minimizer. We are about to present an algorithm for computing global solutions to it.

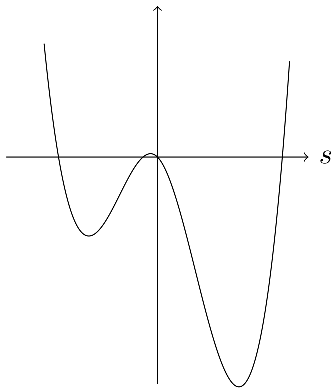

Figure 5.6: Graph of the function $\phi(s) = |s|^3 - 3s^2 - s$. It has one local, nonglobal minimizer.

We derive a solution method for the solution of the cubic subproblem (5.34), where $M > 0$ and H is symmetric. Let us define $\rho = M/6$ and $Q = (1/2)H$. We define

$$\phi(s) = g^T s + s^T Q s + \rho \|s\|_2^3.$$

1. We start computing the eigenvalue decomposition $Q = UDU^T$ of Q. Here $U \in \mathbb{R}^{n\times n}$ is orthogonal and D is a diagonal matrix with the eigenvalues μ_i of D as diagonal entries. Passing from the variables s to $y = U^T s$, we reduce the problem of minimizing ϕ over $s \in \mathbb{R}^n$ to the problem of minimizing ψ defined by

$$\psi(y) = b^T y + \sum_{i=1}^n \mu_i y_i^2 + \rho\Big(\sum_{i=1}^n y_i^2\Big)^{3/2}$$

over $y \in \mathbb{R}^n$, where $b = U^T g$.

2. If y^* is a global minimizer of ψ, then the signs of y_i^* are opposite to those of b_i, that is, $b_i y_i \le 0$. Otherwise, we could reduce the objective function value if we replace y_i^* with $-y_i^*$. Thus minimizing ψ over $y \in \mathbb{R}^n$ is equivalent to minimizing the function

$$\omega(z) = -\sum_{i=1}^n |b_i| z_i + \sum_{i=1}^n \mu_i z_i^2 + \rho\Big(\sum_{i=1}^n z_i^2\Big)^{3/2}$$

over $z \ge 0$. A minimizer z^* of ω over $z \ge 0$ provides us with a global minimizer y^* of ψ via $y_i^* = -\mathrm{sign}(b_i) z_i^*$.

3. Finally, minimizing ω over $z \ge 0$ via substitution of the variables $z_i = \sqrt{\zeta_i}$ reduces to minimizing the function

$$\chi(\zeta) = \sum_{i=1}^n \Big[\mu_i \zeta_i - |b_i|\sqrt{\zeta_i}\Big] + \rho\Big(\sum_{i=1}^n \zeta_i\Big)^{3/2}$$

over $\zeta \geq 0$. Note that the function ζ is convex over $\zeta \geq 0$.

To solve this optimization problem, we rewrite it equivalently as

$$\min_{\zeta, r} \sum_{i=1}^{n} \left[\mu_i \zeta_i - |b_i| \sqrt{\zeta_i} \right] + \rho r^{3/2} \quad \text{s.t.} \quad r \geq 0, \quad \zeta \geq 0, \quad \sum_{i=1}^{n} \zeta_i \leq r.$$

The Lagrange function L of this problem is given by

$$L(\zeta, r, \lambda) = \left[(\mu_i + \lambda) \zeta_i - |b_i| \sqrt{\zeta_i} \right] + \rho r^{3/2} - \lambda r,$$

where $\lambda \in \mathbb{R}$, Note that we "dualize" only the constraint $\sum_{i=0}^{n} \zeta_i \leq r$. Given $\lambda \geq 0$ and assuming $|b_i| > 0$, we can minimize the Lagrange function L over $\lambda \geq 0$ and $r \geq 0$. The minimizers $\zeta(\lambda)$ and $r(\lambda)$ are unique and depend continuously on λ (provided that $|b_i| > 0$). As a result, we can rapidly solve by bisection the Lagrange dual

$$\max_{\lambda \geq 0} \min_{\zeta \geq 0, r \geq 0} L(\zeta, r, \lambda).$$

Given a solution to the dual problem λ^*, we obtain the minimizer of χ via $\zeta^* = \zeta(\lambda^*)$.

5.8 Exercises

Exercise 5.1 (Gradient descent with constant step size I)**.**
Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L \geq 0$. Let f^* be the optimal value of $\min_{x\in\mathbb{R}^n} f(x)$ and suppose that f^* is finite.

We consider the gradient descent method, Algorithm 5.1, with constant step size $\gamma_k = \gamma$.

Show that if $0 < \gamma < (2/L)$, then for all $K \in \mathbb{N}$, the sequence generated by the gradient descent satisfies

$$\min_{0\leq k\leq K} \|\nabla f(x^k)\|_2^2 \leq \frac{1}{\gamma(1 - L\gamma/2)(K+1)} \cdot (f(x^0) - f^*).$$

You can establish your own proof or solve the following subproblems.

1. Show that for all $k \in \mathbb{N}$,
$$f(x^{k+1}) \leq f(x^k) - \gamma(1 - L\gamma/2)\|\nabla f(x^k)\|_2^2.$$
Hint: Use the descent lemma.
2. Show that
$$\gamma(1 - L\gamma/2) \sum_{k=0}^{K} \|\nabla f(x^k)\|_2^2 \leq f(x^0) - f(x^{K+1}).$$
3. Deduce the convergence rate.

Exercise 5.2 (Gradient descent with constant step size II)**.**
Let $f : \mathbb{R}^n \to \mathbb{R}$ be *convex*, differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L > 0$. Let x^* be a solution to $\min_{x\in\mathbb{R}^n} f(x)$.

We consider the gradient descent method, Algorithm 5.1, with constant step size $\gamma_k = \gamma$. Show that if $0 < \gamma < (2/L)$, then for all $k \in \mathbb{N}$,

$$f(x^k) - f(x^*) \leq \frac{\|x^* - x_0\|_2^2}{(k+1)\gamma(2 - L\gamma)}.$$

You can establish your own proof or solve the following subproblems.

1. Define $r_k = \|x^k - x^*\|_2$. Show that
$$r_{k+1}^2 \leq r_k^2 - \gamma(2 - L\gamma)\nabla f(x^k)^T(x^k - x^*).$$
Hint: Review the proof of Theorem 5.7, and use the fact that for all x, $y \in \mathbb{R}^n$,
$$\|\nabla f(x) - \nabla f(y)\|_2^2 \leq L(\nabla f(x) - \nabla f(y))^T(x - y).$$
(This estimate is established in Exercise 5.8.)

2. Show that
$$\gamma(2 - L\gamma)\big[f(x^k) - f(x^*)\big] \le r_k^2 - r_{k+1}^2.$$

3. Show that for all $k \in \mathbb{N}$,
$$f(x^{k+1}) \le f(x^k) - \gamma(1 - L\gamma/2)\|\nabla f(x^k)\|_2^2 \le f(x^k).$$

4. Deduce the convergence rate.

Exercise 5.3 (Gradient descent applied strongly convex quadratic functions)**.**
Let $f : \mathbb{R}^n \to \mathbb{R}$ by defined by $f(x) := (1/2)x^T Ax - b^T x$, where $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite and $b \in \mathbb{R}^n$.

1. Show that $\nabla f(x) = Ax - b$.

2. Show that ∇f is Lipschitz continuous with Lipschitz constant $L = \lambda_{\max}(A)$ (maximum eigenvalue) and that f is strongly convex with parameter $\mu = \lambda_{\min}(A)$ (minimum eigenvalue).

3. Show that $x^* := A^{-1}b$ is the unique minimizer of f.

4. We consider the gradient descent with minimization rule applied to f. Let us define $g_k = Ax^k - b$. Show that if $g_k \neq 0$, then
$$\gamma_k = \frac{g_k^T g_k}{g_k^T A g_k}.$$

5. Consider the gradient method with constant step size $\gamma_k = \gamma > 0$ applied to f. Let L be the maximum eigenvalue of the matrix A. Show that the gradient method with constant step size γ converges to x^* for every starting point x^0 if and only if $\gamma \in (0, 2/L)$.

   Hint: Show that $x^{k+1} - x^* = (I - \gamma A)^{k+1}(x^0 - x^*)$.

Exercise 5.4 (Gradient descent with adaptive step size)**.**
Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L \ge 0$. Let f^* be the optimal value of $\min_{x\in\mathbb{R}^n} f(x)$ and suppose that f^* is finite.

For $x^k \in \mathbb{R}^n$ and $M_k > 0$, we define the model function
$$\phi_{x^k,M_k}(x) := f(x^k) + \nabla f(x^k)^T(x - x^k) + (M_k/2)\|x - x^k\|_2^2.$$
We denote by $x_+^{M_k}$ the minimizer of ϕ_{x^k,M_k}.

We consider the following algorithm for minimizing f.

Algorithm 5.32.

0. Choose initial point/starting point $x^0 \in \mathbb{R}^n$ and $M_{-1} \in (0, L]$.

For $k = 0, 1, 2, \ldots$

1. Compute the smallest nonnegative integer $i_k \in \mathbb{N} \cup \{0\}$ such that with $M_k = 2^{i_k} M_{k-1}$,
$$f(x_+^k) \leq \phi_{x^k, M_k}(x_+^k). \tag{5.37}$$

2. Define $x^{k+1} = x_+^k$ and $M_{k+1} = \max\{M_{-1}, M_k/2\}$.

To analyze Algorithm 5.32, solve the following subproblems.

1. Show that $x_+^k = x^k - (1/M_k)\nabla f(x^k)$.
2. Is Algorithm 5.32 a particular version of the gradient descent method Algorithm 5.1? Can we apply the findings of Exercise 5.1 to analyze the convergence of Algorithm 5.32? How could we describe Algorithm 5.32 in a few words?
3. Show that $M_{-1} \leq M_k \leq 2L$.

   Hint: Use the descent lemma.
4. Show that
$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2M_k}\|\nabla f(x^k)\|_2^2 \leq f(x^k) - \frac{1}{4L}\|\nabla f(x^k)\|_2^2.$$
5. Show that
$$\min_{0 \leq k \leq K} \|\nabla f(x^k)\|_2^2 \leq \frac{4L}{(K+1)} \cdot (f(x^0) - f^*).$$

Exercise 5.5 (Gradient descent with diminishing stepsize).
Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L \geq 0$. Let f^* be the optimal value of $\min_{x \in \mathbb{R}^n} f(x)$ and suppose that f^* is finite.

We consider the gradient descent method, Algorithm 5.1, with diminishing stepsize:
$$\gamma_k > 0, \quad \gamma_k \to 0 \quad \text{as} \quad k \to +\infty, \quad \text{and} \quad \sum_{k=1}^{\infty} \gamma_k = +\infty.$$

Solve the following subproblems.

1. Show that there exists $K \in \mathbb{N}$ and $c > 0$ such that $f(x^{k+1}) \leq f(x^k) - \gamma_k c\|\nabla f(x^k)\|_2^2$ for all $k \geq K$.

2. Deduce that $c\sum_{k=K}^{\infty}\gamma_k\|\nabla f(x^k)\|_2^2 \le f(x^K) - \lim_{k\to\infty} f(x^k) \le f(x^K) - f^*$.
3. Deduce that $\liminf_{k\to\infty}\|\nabla f(x^k)\|_2 = 0$.
4. Show that $\lim_{k\to\infty}\|\nabla f(x^k)\|_2 = 0$.

   Hint: If $\lim_{k\to\infty}\|\nabla f(x^k)\|_2 = 0$ does not hold, then there exists $\varepsilon > 0$ such that $\|\nabla f(x^k)\|_2 \ge 2\varepsilon$ for infinitely many $k \in \mathbb{N}$. Moreover $\liminf_{k\to\infty}\|\nabla f(x^k)\|_2 = 0$ ensures that $\|\nabla f(x^k)\|_2 < \varepsilon$ for infinitely many $k \in \mathbb{N}$. Furthermore use the fact that $\|x^{m+q+1} - x^m\|_2 \le \sum_{k=m}^{q}\gamma_k\|\nabla f(x^k)\|_2$ is valid for all m, $q \in \mathbb{N}$.
5. Deduce that each accumulation point of (x^k) is a stationary point of f.

Exercise 5.6 (Gradient descent with minimization rule)**.**
Establish Theorem 5.3 but with the minimization rule instead of the Armijo rule.

Hint: (i) The decrease of function values resulting from the minimization rule is at least as large as that resulting from the Armijo rule. (ii) The solution to this exercise is short.

Exercise 5.7 (Gradient descent with errors [4]*)**.**
Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L \ge 0$. Let f^* be the optimal value of $\min_{x\in\mathbb{R}^n} f(x)$ and suppose that f^* is finite.

We consider the following algorithm.

Algorithm 5.33.

0. Choose initial point/starting point $x^0 \in \mathbb{R}^n$ and choose

$$\gamma_k > 0 \quad \text{with} \quad \sum_{k=1}^{\infty}\gamma_k = +\infty \quad \text{and} \quad \sum_{k=1}^{\infty}\gamma_k^2 < +\infty.$$

For $k = 0, 1, 2, \ldots$

1. Compute $x^{k+1} = x^k - \gamma_k\nabla f(x^k) + \gamma_k w^k$, where $w^k \in \mathbb{R}^n$.

Suppose that the errors w^k in Algorithm 5.33 satisfies

$$\|w^k\|_2 \le \gamma_k(c_1 + c_2\|\nabla f(x^k)\|_2) \quad \text{for all} \quad k \in \mathbb{N},$$

where c_1 and c_2 are positive constants.

Show that each accumulation point of (x^k) is a stationary point of f.

Exercise 5.8 (Convex functions with Lipschitz continuous gradients)**.**
Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and let $L > 0$. Show that the following statements are equivalent:

1. f is convex and its gradient is Lipschitz continuous with Lipschitz constant L.
2. For all $x, y \in \mathbb{R}^n$,
$$0 \le f(y) - f(x) - \nabla f(x)^T(y-x) \le (L/2)\|y-x\|_2^2.$$
3. For all $x, y \in \mathbb{R}^n$,
$$f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \le f(y).$$
4. For all $x, y \in \mathbb{R}^n$,
$$\frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_2^2 \le (\nabla f(x) - \nabla f(y))^T(x-y).$$

Hints: To show that the second part implies the third, define $g(y) := f(y) - \nabla f(z)^T y$ (with fixed $z \in \mathbb{R}^n$) and use the second part to deduce the estimate
$$g(z) \le g(y) - \frac{1}{2L}\|\nabla g(y)\|_2^2.$$

Exercise 5.9.
Let $f : \mathbb{R}^n \to \mathbb{R}$ be strongly convex with parameter $\mu > 0$, differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L > 0$.
Show that for all $x, y \in \mathbb{R}^n$,
$$\frac{1}{L+\mu}\|\nabla f(x) - \nabla f(y)\|_2^2 + \frac{\mu L}{L+\mu}\|x-y\|_2^2 \le (\nabla f(x) - \nabla f(y))^T(x-y).$$

You can establish your own proof or solve the following subproblems.

1. Show that $L \ge \mu$.

   Hints: Use Exercise 5.8 and Exercise 2.13.
2. Define $\phi(x) := f(x) + (\mu/2)\|x\|_2^2$. Show that for all $x, y \in \mathbb{R}^n$,
$$(\nabla\phi(x) - \nabla\phi(y))^T(x-y) \le (L-\mu)\|x-y\|_2^2.$$

   Hints: Use Exercise 5.8.
3. Deduce the estimate.

Exercise 5.10 (Gradient descent with constant step size III)**.**
Let $f : \mathbb{R}^n \to \mathbb{R}$ be strongly convex with parameter $\mu > 0$, differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L > 0$. Let x^* be a solution to $\min_{x \in \mathbb{R}^n} f(x)$.

We consider the gradient descent method, Algorithm 5.1, with constant step size $\gamma_k = \gamma$. Show that if $0 < \gamma \leq 2/(L + \mu)$, then for all $k \in \mathbb{N}$, the sequence generated by the gradient descent satisfies

$$\|x^k - x^*\|_2^2 \leq \left(1 - \frac{2\gamma\mu L}{\mu + L}\right)^k \|x^0 - x^*\|_2^2.$$

Moreover show that if $\gamma = 2/(L + \mu)$ then

$$\|x^k - x^*\|_2 \leq \left(\frac{\kappa_f - 1}{\kappa_f + 1}\right)^k \|x^0 - x^*\|_2,$$

where $\kappa_f = L/\mu$.

Hints: (i) Show that

$$\|x^{k+1} - x^*\|_2^2 = \|x^k - x^*\|_2^2 - 2\gamma\nabla f(x^k)^T(x^k - x^*) + \gamma^2\|\nabla f(x^k)\|_2^2.$$

(ii) Use Exercise 5.9.

Exercise 5.11 (Gradient descent with constant step size IV)**.**
Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L > 0$. Let f^* be the optimal value of $\min_{x \in \mathbb{R}^n} f(x)$ and suppose that f^* is finite.

Let $0 < \varrho < 2/(2 + L)$. We consider the gradient descent method, Algorithm 5.1, with constant step size γ_k fulfilling

$$\varrho \leq \gamma_k \leq \frac{2(1 - \varrho)}{L} \quad \text{for all} \quad k \in \mathbb{N}.$$

1. Show that for each $k \in \mathbb{N}$,

$$\varrho^2\|\nabla f(x^k)\|_2^2 \leq f(x^k) - f(x^{k+1}).$$

   Hint: Use the descent lemma.

2. Show that for each $k \in \mathbb{N}$,

$$f(x^{k+1}) < f(x^k) \quad \text{provided that} \quad \nabla f(x^k) \neq 0.$$

3. Show that $(f(x^k))$ converges to some finite number $k \to \infty$.

4. Show that $\|\nabla f(x^k)\|_2 \to 0$ as $k \to \infty$.

5. Let $\bar{x}$ be an accumulation point of (x^k). Show that $\nabla f(\bar{x}) = 0$.

Exercise 5.12 (Gradient descent with constant step size V)**.**
We define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) := \|x\|_2^2$. We consider the gradient descent method, Algorithm 5.1, with constant step size $\gamma_k = 1$ and initial point $x^0 := (1/2, 0)$.

1. Show that $\|x^k\|_2 = 1/2$ for all $k \in \mathbb{N}$.

2. Does the previous statement contradict the convergence statement in Exercise 5.11?

Exercise 5.13 (Conditional gradient method for nonconvex optimization)**.**
Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, and let $X \subset \mathbb{R}^n$ be nonempty, closed, bounded, and convex. We denote by $D_X := \sup_{x,y \in X} \|x - y\|_2$ the diameter of X. Let ∇f be Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant L. Suppose that f is bounded from below on X by $f^* \in \mathbb{R}$.

We define the *gap function* gap : $X \to \mathbb{R}$ by

$$\operatorname{gap}(x) := \sup_{v \in X} \nabla f(x)^T (x - v).$$

Algorithm

0. Choose $x^0 \in X$.

For $k = 0, 1, 2, \ldots$:

1. Compute a solution v^k to

$$\min_{v \in X} \nabla f(x^k)^T v.$$

2. Compute γ_k by

$$\gamma_k = \min\left\{1, \frac{\operatorname{gap}(x^k)}{L\|v^k - x^k\|_2^2}\right\} \quad \text{if} \quad v^k \neq x^k, \quad \text{and} \quad \gamma_k = 1 \quad \text{otherwise}.$$

3. Compute $x^{k+1} = \gamma_k v^k + (1 - \gamma_k) x^k$.

Solve the following subproblems.

1. Show that $\operatorname{gap}(x) \geq 0$ for all $x \in X$.

2. Fix $\bar{x} \in X$. Show that $\mathrm{gap}(\bar{x}) = 0$ if and only if
$$\nabla f(\bar{x})^T(x - \bar{x}) \geq 0 \quad \text{for all} \quad x \in X.$$
3. Show that $x^k \in X$ for all $k \in \mathbb{N}$.
4. Show that for all $k \in \mathbb{N}$,
$$f(x^{k+1}) \leq f(x^k) - \gamma_k \mathrm{gap}(x^k) + (L/2)\gamma_k^2 \|x^k - v^k\|_2^2.$$
5. Show that if $\gamma_k = 1$ then
$$f(x^{k+1}) \leq f(x^k) - \mathrm{gap}(x^k)/2.$$
6. Show that if $\gamma_k < 1$ then
$$f(x^{k+1}) \leq f(x^k) - \frac{\mathrm{gap}(x^k)^2}{2L\|x^k - v^k\|_2^2}.$$
7. Show that for all $k \in \mathbb{N}$,
$$\min_{0 \leq \ell \leq k} \mathrm{gap}(x^\ell) \leq \max\left\{\sqrt{\frac{2LD_X^2(f(x^0) - f^*)}{k+1}}, \frac{2(f(x^0) - f^*)}{k+1}\right\}.$$

Exercise 5.14 (proximal gradient method [8]/composite gradient minimization*)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L > 0$. Moreover let $\psi \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, convex and lower semicontinuous.

We develop a prox-gradient-type method for the solution of
$$\min_{x \in \mathbb{R}^n} f(x) + \psi(x).$$
Let $\phi^* \in \mathbb{R}$ be its optimal value. We define $\phi(x) \coloneqq f(x) + \psi(x)$.

Algorithm 5.34.

0. Choose initial point/starting point $x^0 \in \mathrm{dom}(\psi)$, and $\alpha > 0$.

For $k = 0, 1, 2, \ldots$

1. Choose $\alpha_k \geq \alpha$ and compute d^k as the solution to
$$\min_{d \in \mathbb{R}^n} \nabla f(x^k)^T d + (\alpha_k/2)\|d - x^k\|_2^2 + \psi(d).$$
2. Define $s^k \coloneqq d^k - x^k$.

3. Choose a step size $\gamma_k \in [0,1]$ and define $x^{k+1} = x^k + \gamma_k s^k$.

In order to analyze Algorithm 5.34, solve the following subproblems.

1. Show that Algorithm 5.34 is well-defined.
2. Show that for all $0 < \gamma \leq 1$,
$$\frac{\psi(x^k + \gamma s^k) - \psi(x^k)}{\gamma} \leq \psi(x^k + s^k) - \psi(x^k).$$
Hint: Use Exercise 2.28.
3. Show that
$$\phi'(x^k; s^k) \leq \nabla f(x^k)^T s^k + \psi(s^k + x^k) - \psi(x^k).$$
Hints: Use the identity $\phi'(x^k; s^k) = \nabla f(x^k)^T s^k + \psi'(x^k; s^k)$ and use Exercise 2.28.
4. Show that
$$\nabla f(x^k)^T s^k + \psi(s^k + x^k) - \psi(x^k) \leq -\alpha_k \|s^k\|_2^2.$$
Hints: Use the identity $d^k = s^k + x^k$ and the optimality conditions derived in Exercise 2.25.
5. Let $0 < \gamma \leq 1$. Show that
$$\phi(x^k + \gamma s^k) \leq \phi(x^k) + ((L/2)\gamma^2 - \alpha\gamma)\|s^k\|_2^2.$$
6. We define $\gamma^* := \min\{1, \alpha/L\}$. Let $\gamma_k = \gamma^*$ for all $k \in \mathbb{N}$. Show that
$$\min_{1 \leq k \leq K} \|s^k\|_2^2 \leq \frac{2L}{\alpha} \cdot \frac{1}{K+1} \cdot (\phi(x^0) - \phi^*).$$
7. For a convex, proper function $g \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, we define its prox-mapping by
$$\operatorname{prox}_g(x) := \operatorname{argmin}_{y \in \mathbb{R}^n} (1/2)\|y - x\|_2^2 + g(y).$$
Show that
$$s^k = \operatorname{prox}_{(1/\alpha_k)\psi}\big(x^k - (1/\alpha_k)\nabla f(x^k)\big) - x^k.$$
(Compare with Exercise 2.26).
8. Show that if $\psi = 0$ then $s^k = -(1/\alpha_k)\nabla f(x^k)$.

Exercise 5.15 (Can a gradient method with summable step sizes converge?)**.**
Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, let $x^0 \in \mathbb{R}^n$, and let $S := \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ be bounded. Denote by C the optimal value of $\max_{x \in S} \|\nabla f(x)\|_2$. (Why is C finite?) We consider the algorithm $x^{k+1} = x^k - \gamma_k \nabla f(x^k)$, where $\gamma_k \geq 0$. Suppose that $(x^k) \subset S$ and that $\Gamma := \sum_{k=0}^{\infty} \gamma_k < \infty$. Let $\bar{x}$ be a stationary point of f.

Note: (i) The results of this exercise have been used to construct Example 5.2. (ii) $(x^k) \subset S$ is true if $f(x^{k+1}) \leq f(x^k)$ for all $k \in \mathbb{N} \cup \{0\}$, for example. (iii) $\sum_{k=0}^{\infty} \gamma_k < \infty$ is true if $\gamma_k := c/(1+k)^2$, where $c > 0$ is a constant, for example.

1. Show that for all $k \in \mathbb{N}$,
$$\|x^{k+1} - x^0\|_2 \leq C\Gamma.$$
This means that $(x^k) \subset \mathrm{cl}(B_{C\Gamma}(x^0))$.
Hint: Use the fact that $x^{k+1} - x^0 = \sum_{i=0}^{k} (x^{i+1} - x^i)$.
Note: We have $\mathrm{cl}(B_{C\Gamma}(x^0)) = \{x \in \mathbb{R}^n : \|x - x^0\|_2 \leq C\Gamma\}$ (the closed ball about x^0 with radius $C\Gamma$).

2. Show that
$$\|x^{k+1} - \bar{x}\|_2 \geq \|\bar{x} - x^0\|_2 - C\Gamma.$$

3. Deduce that if $\|\bar{x} - x^0\|_2 > C\Gamma$ then
$$\liminf_{k\to\infty} \|x^{k+1} - \bar{x}\|_2 > 0.$$

Exercise 5.16 (Successive stepsize reduction rule without convergence [2, Figure 1.2.6])**.**
We appy the iterative scheme $x^{k+1} = x^k - \gamma_k \nabla f(x^k)$ to

$$f(x) = \begin{cases} \frac{3(1-x)^2}{4} - 2(1-x) & \text{if} \quad x > 1, \\ \frac{3(1+x)^2}{4} - 2(1+x) & \text{if} \quad x < -1, \\ x^2 - 1 & \text{if} \quad -1 \leq x \leq 1. \end{cases}$$

Let $\beta \in (0,1)$. We compute γ_k according to the following heuristic stepsize reduction rule: Compute the largest $\gamma_k \in \{1, \beta, \beta^2, \ldots\}$ such that

$$f(x^k - \gamma_k \nabla f(x^k)) < f(x^k).$$

Let $|x^0| > 1$.

1. How does the heuristic stepsize reduction rule differ from backtracking linesearch?

2. Show that $|x^k| > 1$ for all $k \in \mathbb{N}$.
Hints:

a) Show that if $x > 1$ then $x - \nabla f(x) < -1$ and $f(x - \nabla f(x)) < f(x)$.

b) Show that if $x < -1$ then $x - \nabla f(x) > 1$ and $f(x - \nabla f(x)) < f(x)$.

3. Show that f is continuously differentiable and strictly convex, and show that $x^* := 0$ is the unique minimizer of f.

4. Show that no subsequence of (x^k) converges to the stationary point of f. Does this statement contradict Theorem 5.3.

Exercise 5.17 (Newton proximal extragradient (NPE) method [12, section 7])**.**
Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be monotone (that is, $(F(x) - F(y))^T(x - y) \geq 0$ for all $x,\, y \in \mathbb{R}^n$), and differentiable. Let F' (the Jacobian of F) be Lipschitz continuous with Lipschitz constant $L > 0$. Let $x^* \in \mathbb{R}^n$ satisfy $F(x^*) = 0$.

Algorithm 5.35 (Newton proximal extragradient (NPE) method)**.**

0. Let $x^0 \in \mathbb{R}^n$, and let $0 < \sigma_\ell < \sigma_u < 1$.

For $k = 1, 2, \ldots$

1. Compute $\lambda_k > 0$ and $s^k \in \mathbb{R}^n$ such that

$$(F'(x^{k-1}) + (1/\lambda_k)I)s^k = -F(x^{k-1}), \quad \text{and} \quad \frac{2}{L}\sigma_\ell \leq \lambda_k \|s^k\|_2 \leq \frac{2}{L}\sigma_u.$$

2. Define $y^k = x^{k-1} + s^k$ and $x^k = x^{k-1} - \lambda_k F(y^k)$.

Show that for all $K \in \mathbb{N}$,

$$\min_{1 \leq k \leq K} \|F(y^k)\|_2 \leq \frac{L(1 + 1/\sigma_\ell)\|x^* - x^0\|_2^2}{2(1 - \sigma_u^2)} \cdot \frac{1}{K}.$$

You can establish your own proof or solve the following subproblems.

1. Show that $F'(x)$ is positive semidefinite for all $x \in \mathbb{R}^n$. Deduce that the algorithm is well-defined.

2. Show that

$$\|F(y^k)\|_2 \leq (L/2)\|s^k\|_2^2 + (1/\lambda_k)\|s^k\|_2 \leq (L/2)(1 + 1/\sigma_\ell)\|s^k\|_2^2.$$

Hints: (i) For all $x,\, s \in \mathbb{R}^n$, we have

$$\|F(x + s) - F(x) - F'(x)s\|_2 \leq (L/2)\|s\|_2^2.$$

This holds true because F' is L-Lipschitz continuous. (ii) Review the proof of Theorem 5.31.

3. Show that $F'(y^k)^T(y^k - x^*) \geq 0$.

4. Show that

$$\|y^k - x^k\|_2 = \|\lambda_k F(y^k) + y^k - x^{k-1}\|_2 \leq (L\lambda_k/2)\|s^k\|_2^2 \leq \sigma_u \|s^k\|_2.$$

   Deduce that

$$(1 - \sigma_u)\|s^k\|_2 \leq \lambda_k \|F(y^k)\|_2 \leq (1 + \sigma_u)\|s^k\|_2.$$

   Hint: Define $G_k(x) := \lambda_k F(x) + x - x^{k-1}$. Show that G_k' is $\lambda_k L$-Lipschitz continuous and that $G_k'(x_{k-1})s^k + G_k(x_{k-1}) = 0$.

5. We define $r_k := \|x^k - x^*\|_2$. Show that

$$r_k^2 = r_{k-1}^2 + 2(x^* - y^k)^T(x^{k-1} - x^k) + \|y^k - x^k\|_2^2 - \|y^k - x^{k-1}\|_2^2.$$

6. Deduce that

$$r_k^2 \leq r_{k-1}^2 - (1 - \sigma_u^2)\|s^k\|_2^2.$$

   Deduce that $\|x^k - x^*\|_2 \leq \|x^{k-1} - x^*\|_2$.

7. Deduce that $(1 - \sigma_u^2)\sum_{k=1}^K \|s^k\|_2^2 \leq \|x^* - x^0\|_2^2$.

8. Deduce the convergence rate.

   Hint: Use part two.

6 Optimization Methods for Constrained Optimization

Materials will be added at a later point in time.

Bibliography

[1] A. Ben-Tal and A. Nemirovski. Optimization III: Convex analysis, nonlinear programming theory, standard nonlinear programming algorithms. lecture notes, Georgia Institute of Technology, 2023. URL: https://www2.isye.gatech.edu/~nemirovs/.

[2] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, MA, 2003.

[3] D. P. Bertsekas. *Convex Optimization Theory.* Athena Scientific, Nashua, NH, 2009. Solution manual available at http://www.athenasc.com/convexduality.html. URL: http://web.mit.edu/dimitrib/www/Convex_Theory_Entire_Book.pdf.

[4] D. P. Bertsekas and J. N. Tsitsiklis. Gradient convergence in gradient methods with errors. *SIAM J. Optim.*, 10(3):627–642, 2000. doi:10.1137/S1052623497331063.

[5] A. A. Borovkov. *Asymptotic analysis of random walks: Light-tailed distributions*, volume 176 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 2020. Translated by V. V. Ulyanov and M. V. Zhitlukhin. doi:10.1017/9781139871303.

[6] A. A. Borovkov and A. A. Mogulskii. Chebyshev-type exponential inequalities for sums of random vectors and for trajectories of random walks. *Theory Probab. Appl.*, 56(1):21–43, 2012. doi:10.1137/S0040585X97985182.

[7] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, Cambridge, 2004. URL: https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf, doi:10.1017/CBO9780511804441.

[8] M. Fukushima and H. Mine. A generalized proximal point algorithm for certain nonconvex minimization problems. *Internat. J. Systems Sci.*, 12(8):989–1000, 1981. doi:10.1080/00207728108963798.

[9] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 2nd edition, 2013. doi:10.1017/CBO9781139020411.

[10] G. Lan. Lectures for convex and nonlinear optimization. lecture notes, Georgia Institute of Technology, 2022.

[11] G. Lan. *First-order and Stochastic Optimization Methods for Machine Learning.* Springer Ser. Data Sci. Springer, Cham, 2020. doi:10.1007/978-3-030-39568-1.

[12] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM J. Optim.*, 20(6):2755–2787, 2010. `doi:10.1137/090753127`.

[13] Yu. Nesterov. *Lectures on Convex Optimization.* Springer Optim. Appl. 137. Springer, Cham, 2nd edition, 2018. `doi:10.1007/978-3-319-91578-4`.

[14] H. Sun, A. Shapiro, and X. Chen. Distributionally robust stochastic variational inequalities. *Math. Program.*, 200(1):279–317, 2023. `doi:10.1007/s10107-022-01889-2`.

[15] N. J. Walkington. Nesterov's method for convex optimization. *SIAM Rev.*, 65(2):539–562, 2023. `doi:10.1137/21M1390037`.

Lecture minutes

December 5, 2023

- Covered slides of chapter 6.

November 30, 2023

- Covered slides of chapter 5 up to slide 55.
- Proved part 3 of the proposition on slide 53.
- Proved theorem on slide 55.

November 28, 2023

- Covered slides of chapter 5 up to slide 53.
- Proved parts 1–2 of the proposition on slide 53.

November 21, 2023

- Covered slides of chapter 5 up to slide 47.
- Solvers for nonlinear optimization such as IPOPT, SNOPT, KNITRO support L-BFGS.
- No proofs.

November 16, 2023

- Covered slides of chapter 5 up to slide 36.
- Compared q-superlinear and q-quadratic convergence on example sequences
- Proved the theorem on page 32.

November 14, 2023

- Covered *updated* slides of chapter 5 up to slide 31.
- Compared empirical performance of AGD and GD on an example problem.
- Discussed example *"Newton's method may fail to converge"* provided on slide 35.
- No proofs.

Solutions to Homework

1 Homework 1

Exercise 1.1.
Which of the following sets are convex?

1. $\{x \in \mathbb{R}^2 \colon x_1 + i^2 x_2 \le 1,\ i = 1, \dots, 10\}$
2. $\{x \in \mathbb{R}^2 \colon x_1^2 + 2ix_1x_2 + i^2x_2^2 \le 1,\ i = 1, \dots, 10\}$
3. $\{x \in \mathbb{R}^2 \colon x_1^2 + ix_1x_2 + i^2x_2^2 \le 1,\ i = 1, \dots, 10\}$
4. $\{x \in \mathbb{R}^2 \colon x_1^2 + 5x_1x_2 + 4x_2^2 \le 1\}$

Solution:

1. $\{x \in \mathbb{R}^2 \colon x_1 + i^2 x_2 \le 1,\ i = 1, \dots, 10\}$ is convex. It is a polyhedral set. Defining $a_i = (0, i^2)$ for $i = 1, \dots, 10$, we obtain $\{x \in \mathbb{R}^2 \colon x_1 + i^2 x_2 \le 1,\ i = 1, \dots, 10\} = \{x \in \mathbb{R}^2 \colon a_i^T x \le 1,\ i = 1, \dots, 10\}$.
2. $\{x \in \mathbb{R}^2 \colon x_1^2 + 2ix_1x_2 + i^2x_2^2 \le 1,\ i = 1, \dots, 10\}$ is convex. We have $x_1^2 + 2ix_1x_2 + i^2x_2^2 = (x_1 + ix_2)^2$. So the set is the intersection of the sets $\{x \in \mathbb{R}^n \colon |x_1 + ix_2| \le 1\}$ for $i = 1, \dots, 10$. Moreover, $\{x \in \mathbb{R}^n \colon |x_1 + ix_2| \le 1\}$ is polyhedral for each $i \in \{1, \dots, 10\}$.
3. $\{x \in \mathbb{R}^2 \colon x_1^2 + ix_1x_2 + i^2x_2^2 \le 1,\ i = 1, \dots, 10\}$ is convex, as it is the intersection of ellipsoids. Fix $i \in \{1, \dots, 10\}$. We define

   $$Q_i = \begin{bmatrix} 1 & i/2 \\ i/2 & i^2 \end{bmatrix}.$$

   The matrix Q_i is symmetric and positive definite. The latter fact can be verified using Sylvester's criterion (the determinate of Q_i is $i^2 - (i/2)^2 = i^2 - i^2/4 > 0$). We have

   $$x^T Q_i x = x^T \begin{bmatrix} x_1 + (i/2)x_2 \\ (i/2)x_1 + i^2 x_2 \end{bmatrix} = x_1^2 + (i/2)x_1x_2 + (i/2)x_1x_2 + i^2x_2^2.$$

   Hence $x_1^2 + ix_1x_2 + i^2x_2^2 \le 1$ if and only if $x^T Q_i x \le 1$.
4. $\{x \in \mathbb{R}^2 \colon x_1^2 + 5x_1x_2 + 4x_2^2 \le 1\}$ is nonconvex. This can be deduced from a visualization of the set, for example. For instance, WolframAlpha can be used to visualize this set. Click here to see a WolframAlpha plot.

Exercise 1.2 (see [7, Exercise 2.2])**.**
Show that a set is convex if and only if its intersection with any line is convex.

Solution: The intersection of two convex sets is convex. Therefore if S is a convex set, the intersection of S with a line is convex.

Conversely, suppose the intersection of S with any line is convex. Take any two distinct points x and $y \in S$. The intersection of S with the line through x and y is convex. The convex combination of x and y belong to the intersection, hence also to S.

Exercise 1.3 (Expanded and restricted sets, see [7, Exercise 2.14])**.**
Let $S \subset \mathbb{R}^n$, and let $\|\cdot\|$ be a norm on $\mathbb{R}^n$.

1. For $a \geq 0$, we define $S_a = \{x \in \mathbb{R}^n : \operatorname{dist}(x, S) \leq a\}$, where $\operatorname{dist}(x, S) = \inf_{y \in S} \|x - y\|$. We refer to S_a as S *expanded* or *extended* by a. Show that if S is nonempty and convex, then S_a is convex.
2. For $a \geq 0$, we define $S_{-a} = \{x \in \mathbb{R}^n : B(x, a) \subset S\}$, where $B(x, a)$ is the closed ball (in the norm $\|\cdot\|$) centered at x with radius a, that is, $B(x, a) = \{y \in \mathbb{R}^n : \|y - x\| \leq a\}$. We refer to S_{-a} as S *shrunk* or *restricted* by a, since S_{-a} consists of all points that are at least a distance a from $\mathbb{R}^n \setminus S$. Show that if S is convex, then S_{-a} is convex.

Solution:

1. To show this convexity statement, we observe at $x \in S_a$ if and only if for each $a' > a$, there exists $y \in S$ such that $\|x - y\| \leq a'$. (This fact is a consequence of the definition of the infimum, $S \neq \emptyset$, and $a < \infty$.) Now, let x, $y \in S_a$ and let $\lambda \in [0, 1]$. We fix $a' > a$. The above statement implies that there exist u, $v \in S$ with $\|x - u\| \leq a'$ and $\|y - v\| \leq a'$. Defining $w = \lambda u + (1 - \lambda)v$, we find that $w \in S$ and

$$\begin{aligned}\operatorname{dist}(\lambda x + (1 - \lambda)y, S) &\leq \|\lambda x + (1 - \lambda)y - w\| \\ &= \|\lambda(x - u) + (1 - \lambda)(y - v)\| \\ &\leq \lambda\|x - u\| + (1 - \lambda)\|y - v\| \leq a'.\end{aligned}$$

   Since $a' > a$ is arbitrary, we find that $\operatorname{dist}(\lambda x + (1 - \lambda)y, S) \leq a$. Hence S_a is a convex set.

2. Consider two points x, $y \in S_{-a}$. Hence for all u with $\|u\| \leq a$, we have

$$x + u \in S, \quad y + u \in S.$$

   For $\lambda \in [0, 1]$ and $\|u\| \leq a$,

$$\lambda x + (1 - \lambda)y + u = \lambda(x + u) + (1 - \lambda)(y + u) \in S,$$

   because S is convex. We conclude that $\lambda x + (1 - \lambda)y \in S_{-a}$.

Exercise 1.4 (A set of hyperplanes, see [7, Exercise 2.21]).
Suppose that C and D are disjoint sets in $\mathbb{R}^n$. Consider the set of points $(a,b) \in \mathbb{R}^{n+1}$ for which $a^T x \le b$ for all $x \in C$ and $a^T x \ge b$ for all $x \in D$. Show that this set is a cone.
Hint: Use Example 1.5.

Solution: The conditions $a^T x \le b$ for all $x \in C$ and $a^T x \ge b$ for all $x \in D$ form a set of homogeneous linear inequalities in (a,b). Therefore, this set is the intersection of halfspaces that pass through the origin. Hence it is a convex cone.

We use provide some further details. We have to show that the set

$$\{(a,b)\colon a^T x \le b, \text{ for all } x \in C, \ a^T x \ge b \text{ for all } x \in D\}$$

is a cone. We claim that it is a solution set of an infinite homogeneous system of nonstrict linear inequalities which is a cone according to Example 1.5.

Indeed, we have

$$\begin{aligned}&\{(a,b)\colon a^T x \le b, \text{ for all } x \in C, \ a^T x \ge b \text{ for all } x \in D\}\\ &\quad = \{(a,b)\colon (x,-1)^T(a,b) \le 0, \text{ for all } x \in C, \ (x,-1)^T(a,b) \ge 0 \text{ for all } x \in D\}.\end{aligned}$$

Here $(x,-1) \in \mathbb{R}^n$ is the vector x augmented by the additional component -1.

Exercise 1.5 (see [7, Exercise 2.12]).
Which of the following sets are convex?

1. A slap, that is, a set of the form $\{x \in \mathbb{R}^n\colon \alpha \le a^T x \le \beta\}$.
2. A rectangle, that is, a set of the form $\{x \in \mathbb{R}^n\colon \alpha_i \le x_i \le \beta_i, \ i = 1,\dots,n\}$.
3. A wedge, that is, $\{x \in \mathbb{R}^n\colon a_1^T x \le b_1, \ a_2^T x \le b_2\}$.
4. The set of points closer to a given point than a given set, that is,
$$\{x \in \mathbb{R}^n\colon \|x - x_0\|_2 \le \|x - y\|_2 \quad \text{for all} \quad y \in S\},$$
where $S \subset \mathbb{R}^n$.

Solution:

1. The set is convex; it can be written as the intersection of $\{x \in \mathbb{R}^n\colon \alpha \le a^T x\}$ and $\{x \in \mathbb{R}^n\colon a^T x \le \beta\}$.
2. The set is convex and can be written as the intersection of $2n$ convex sets.
3. A wedge is the intersection of two convex sets and hence convex.
4. The set is convex. It can be expressed as
$$\bigcap_{y \in S} \{x \in \mathbb{R}^n\colon \|x - x_0\|_2 \le \|x - y\|_2\}.$$
For each $y \in S$, the set $\{x \in \mathbb{R}^n\colon \|x - x_0\|_2 \le \|x - y\|_2\}$ is polyhedral. We have
$$\|x - x_0\|_2^2 = x^T x - 2x^T x_0 + x_0^T x_0 \quad \text{and} \quad \|x - y\|_2^2 = x^T x - 2y^T x + y^T y.$$
Hence
$$\{x \in \mathbb{R}^n\colon \|x - x_0\|_2 \le \|x - y\|_2\} = \{x \in \mathbb{R}^n\colon -2x^T(x_0 - y) + x_0^T x_0 \le y^T y\}.$$
The computations show that $\{x \in \mathbb{R}^n\colon \|x - x_0\|_2 \le \|x - y\|_2\}$ is polyhedral.

2 Homework 2

Exercise 1.1.
Which of the following sets are convex? No justifications are required.

1. $\{x \in \mathbb{R}^2 \colon x_1 + i^2 x_2 \leq 1,\ i = 1, \dots, 10\}$
2. $\{x \in \mathbb{R}^2 \colon x_1^2 + 2ix_1x_2 + i^2x_2^2 \leq 1,\ i = 1, \dots, 10\}$
3. $\{x \in \mathbb{R}^2 \colon x_1^2 + ix_1x_2 + i^2x_2^2 \leq 1,\ i = 1, \dots, 10\}$
4. $\{x \in \mathbb{R}^2 \colon x_1^2 + 5x_1x_2 + 4x_2^2 \leq 1\}$
5. $\{x \in \mathbb{R}^{10} \colon x_1^2+2x_2^2+\dots+10x_{10}^2 \leq 2004x_1-2003x_2+2002x_3-\dots+1996x_9-1995x_{10}\}$
6. $\{x \in \mathbb{R}^2 \colon \exp(x_1) \leq x_2\}$
7. $\{x \in \mathbb{R}^2 \colon \exp(x_1) \geq x_2\}$
8. $\{x \in \mathbb{R}^n \colon \sum_{i=1}^n x_i^2 = 1\}$
9. $\{x \in \mathbb{R}^n \colon \sum_{i=1}^n x_i^2 \leq 1\}$
10. $\{x \in \mathbb{R}^n \colon \sum_{i=1}^n x_i^2 \geq 1\}$
11. $\{x \in \mathbb{R}^n \colon \max_{1\leq i\leq n} x_i \leq 1\}$
12. $\{x \in \mathbb{R}^n \colon \max_{1\leq i\leq n} x_i \geq 1\}$
13. $\{x \in \mathbb{R}^n \colon \max_{1\leq i\leq n} x_i = 1\}$
14. $\{x \in \mathbb{R}^n \colon \min_{1\leq i\leq n} x_i \leq 1\}$
15. $\{x \in \mathbb{R}^n \colon \min_{1\leq i\leq n} x_i \geq 1\}$
16. $\{x \in \mathbb{R}^n \colon \min_{1\leq i\leq n} x_i = 1\}$

Solution:

1. $\{x \in \mathbb{R}^2 \colon x_1 + i^2 x_2 \leq 1,\ i = 1, \dots, 10\}$ is convex. It is a polyhedral set. Defining $a_i = (0, i^2)$ for $i = 1, \dots, 10$, we obtain $\{x \in \mathbb{R}^2 \colon x_1 + i^2x_2 \leq 1,\ i = 1, \dots, 10\} = \{x \in \mathbb{R}^2 \colon a_i^T x \leq 1,\ i = 1, \dots, 10\}$.
2. $\{x \in \mathbb{R}^2 \colon x_1^2 + 2ix_1x_2 + i^2x_2^2 \leq 1,\ i = 1, \dots, 10\}$ is convex. We have $x_1^2 + 2ix_1x_2 + i^2x_2^2 = (x_1+ix_2)^2$. So the set is the intersection of the sets $\{x \in \mathbb{R}^n \colon |x_1+ix_2| \leq 1\}$ for $i = 1, \dots, 10$. Moreover, $\{x \in \mathbb{R}^n \colon |x_1 + ix_2| \leq 1\}$ is polyhedral for each $i \in \{1, \dots, 10\}$.

3. $\{x \in \mathbb{R}^2 \colon x_1^2 + ix_1x_2 + i^2x_2^2 \le 1,\ i = 1, \dots, 10\}$ is convex, as it is the intersection of ellipsoids. Fix $i \in \{1, \dots, 10\}$. We define

$$Q_i = \begin{bmatrix} 1 & i/2 \\ i/2 & i^2 \end{bmatrix}.$$

The matrix Q_i is symmetric and positive definite. The latter fact can be verified using Sylvester's criterion (the determinate of Q_i is $i^2 - (i/2)^2 = i^2 - i^2/4 > 0$). We have

$$x^T Q_i x = x^T \begin{bmatrix} x_1 + (i/2)x_2 \\ (i/2)x_1 + i^2x_2 \end{bmatrix} = x_1^2 + (i/2)x_1x_2 + (i/2)x_1x_2 + i^2x_2^2.$$

Hence $x_1^2 + ix_1x_2 + i^2x_2^2 \le 1$ if and only if $x^T Q_i x \le 1$.

4. $\{x \in \mathbb{R}^2 \colon x_1^2 + 5x_1x_2 + 4x_2^2 \le 1\}$ is nonconvex. This can be deduced from a visualization of the set, for example. For instance, WolframAlpha can be used to visualize this set. Click here to see a WolframAlpha plot.

5. $\{x \in \mathbb{R}^{10} \colon x_1^2 + 2x_2^2 + \cdots + 10x_{10}^2 \le 2004x_1 - 2003x_2 + 2002x_3 - \cdots + 1996x_9 - 1995x_{10}\}$ is convex, as it is an ellipsoid.

6. $\{x \in \mathbb{R}^2 \colon \exp(x_1) \le x_2\}$ is convex. Click here for a graphical illustration.

7. $\{x \in \mathbb{R}^2 \colon \exp(x_1) \ge x_2\}$ is nonconvex.

8. $\{x \in \mathbb{R}^n \colon \sum_{i=1}^n x_i^2 = 1\}$ is nonconvex. This set is the boundary of the closed Euclidean unit norm ball.

9. $\{x \in \mathbb{R}^n \colon \sum_{i=1}^n x_i^2 \le 1\}$ is convex. It represents the closed Euclidean unit norm ball in $\mathbb{R}^n$.

12. $X \coloneqq \{x \in \mathbb{R}^n \colon \max_{1 \le i \le n} x_i \ge 1\}$ is nonconvex if $n > 1$ and convex if $n = 1$. If $n = 1$, then the set is the interval $[1, +\infty)$. Now let $n > 1$. We define $x = (1, 0, \dots, 0)$ and $y = (0, 1, 0, \dots, 0)$. We have $x, y \in X$ and $(1/2)x + (1/2)y \notin X$. See for example here for a graphical illustration for $n = 2$.

13. $\{x \in \mathbb{R}^n \colon \max_{1 \le i \le n} x_i = 1\}$ is nonconvex if $n > 1$ and convex if $n = 1$.

14. $\{x \in \mathbb{R}^n \colon \min_{1 \le i \le n} x_i \le 1\}$ is nonconvex if $n > 1$ and convex if $n = 1$.

15. $\{x \in \mathbb{R}^n \colon \min_{1 \le i \le n} x_i \ge 1\}$ is convex. It is the intersection of convex sets because

$$\begin{aligned} \{x \in \mathbb{R}^n \colon \min_{1 \le i \le n} x_i \ge 1\} &= \{x \in \mathbb{R}^n \colon x_i \ge 1, \quad i = 1, \dots, n\} \\ &= \bigcap_{i=1,\dots,n} \{x \in \mathbb{R}^n \colon x_i \ge 1\} \end{aligned}$$

16. $\{x \in \mathbb{R}^n \colon \min_{1 \le i \le n} x_i = 1\}$

Exercise 1.6.
Which of the following statements are true? For true statements, provide proofs; for wrong statements, provide counterexamples.

1. The convex hull of a closed set in $\mathbb{R}^n$ is closed.

2. The convex hull of a closed convex set in $\mathbb{R}^n$ is closed.

3. The convex hull of a closed, bounded set in $\mathbb{R}^n$ is closed and bounded.

Hints: (i) Use the fact that a set in $\mathbb{R}^n$ is compact if and only if it is bounded and closed. (ii) Continuous functions map compact sets to compact ones.

Solution:

1. The convex hull of a closed set in $\mathbb{R}^n$ may not be closed. For example, the convex hull of the closed set $X := \{x \in \mathbb{R}^2 \colon x_2 \geq |x_1|^{-1},\ x_1 \neq 0\}$ is the open halfplane $\{x \in \mathbb{R}^2 \colon x_2 > 0\}$. Click here for a graphical illustration of X.

2. The convex hull of a closed convex set in $\mathbb{R}^n$ equals the closed convex set and hence is closed. In other words, if $X \subset \mathbb{R}^n$ is closed and convex, then $\mathrm{Conv}(X) = X = \mathrm{cl}(X)$. Hence $\mathrm{Conv}(X)$ is closed.

3. This statement is true. By Caratheodory's theorem, the convex hull of a set $X \subset \mathbb{R}^n$ is the set of all vectors $\sum_{i=1}^{n+1} \lambda_i x^i$ with $x^i \in X$, $\lambda_i \geq 0$, and $\sum_{i=1}^{n+1} \lambda_i = 1$. Let us define the set $\Lambda = \{\lambda \in \mathbb{R}^{n+1} \colon \lambda_i \geq 0,\ \sum_{i=1}^{n+1} \lambda_i = 1\}$. The convex hull of X is the image of the set

$$\Lambda \times \underbrace{X \times \cdots \times X}_{\text{taken } n+1 \text{ times}}$$

under the continuous mapping $(\lambda, x^1, \ldots, x^{n+1}) \mapsto \sum_{i=1}^{n+1} \lambda_i x^i$. If X is compact, then $\Lambda \times X \times \cdots \times X$ is a compact set and hence the image of this set under the above continuous mapping is compact. Putting together the pieces, we obtain the assertion.

Exercise 1.11.
Let $A \in \mathbb{R}^{n \times n}$ be symmetric positive definite and let $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Show that if $b^T A^{-1} b - c > 0$, then the set

$$X = \{x \in \mathbb{R}^n \colon x^T A x + 2b^T x + c \leq 0\}$$

is an ellipsoid. (If $c \leq 0$ and $b \neq 0$ or if $c < 0$ and $b = 0$, then $b^T A^{-1} b - c > 0$).

Solution: We have to show that we can represented X by

$$X = \{x \in \mathbb{R}^n \colon (x-a)^T Q(x-a) \leq r^2\},$$

where $Q \in \mathbb{R}^{n\times n}$ is symmetric positive definite, $a \in \mathbb{R}^n$, and $r > 0$.

We compute

$$\begin{aligned}(x-a)^T Q(x-a) &= x^T Q(x-a) - a^T Q(x-a)\\ &= x^T Qx - a^T Qx - a^T Qx + a^T Qa\\ &= x^T Qx - 2a^T Qx + a^T Qa.\end{aligned}$$

Comparing both representations, we identify $Q = A$, $b = -Qa$, and $r^2 - a^T Qa = -c$. We choose $Q = A$, $a = -A^{-1}b$, and $r = \sqrt{a^T Aa - c} = \sqrt{b^T A^{-1}b - c}$. By assumption, we have $r > 0$.

Note that if $x = -A^{-1}b \in X$, then we have $b^T A^{-1}b - c \geq 0$ because

$$0 \geq x^T Ax + 2b^T x + c = b^T A^{-1}b - 2b^T A^{-1}b + c = -b^T A^{-1}b + c.$$

Exercise 1.12 (Kirchberger's Theorem)**.**
Prove the following Kirchberger's Theorem:

Let $X = \{x^1, \ldots, x^k\}$ and $Y = \{y^1, \ldots, y^m\}$ be finite subsets in $\mathbb{R}^n$ with and $k+m \geq n+2$. Let the points $x^1, \ldots, x^k$, $y^1, \ldots, y^m$ be distinct. If for each subset $S \subset X \cup Y$ with $n+2$ points the convex hulls of the sets $X \cap S$ and $Y \cap S$ are disjoint, then convex hulls of X and Y are disjoint.

Hint: Either one of the following approaches might be helpful:

1. Assume on the contrary that the convex hulls of X and Y intersect so that

$$\sum_{i=1}^{k} \lambda_i x^i = \sum_{j=1}^{m} \mu_j y^j$$

   for certain nonnegative numbers $\lambda_i \geq 0$, $\mu_j \geq 0$ with $\sum_{i=1}^k \lambda_i = 1$ and $\sum_{j=1}^p \mu_j = 1$, and look at this expression with the minimum total number of nonzero coefficients λ_i, μ_j.

2. Show that if the convex hulls of X and Y intersect, then there exists a set $T \subset X \cup Y$ with at most $n+2$ elements such that the convex hulls of the sets $X \cap T$ and $Y \cap T$ intersect. Subsequently, deduce Kirchberger's theorem.

Solution: We use the second approach. Let $z \in \operatorname{Conv}(X)$ and $z \in \operatorname{Conv}(Y)$. Then there exist nonnegative numbers $\lambda_i \geq 0$, $\mu_j \geq 0$ with $\sum_{i=1}^k \lambda_i = 1$ and $\sum_{j=1}^p \mu_j = 1$ such that

$$z = \sum_{i=1}^{k} \lambda_i x^i = \sum_{j=1}^{m} \mu_j y^j$$

Let p be the number of λ_i's with $\lambda_i > 0$ and and let q be the number of μ_j's with $\mu_j > 0$. We assume without loss of generality that $\lambda_1, \dots, \lambda_p > 0$ and $\lambda_{p+1} = \cdots = \lambda_k = 0$ and $\mu_1, \dots, \mu_q > 0$ and $\mu_{q+1} = \cdots = \mu_m = 0$.

We consider the set $\widehat{T} = \{x^1, \dots, x^p, y^1, \dots, y^q\}$. This set has $p+q$ elements, and $\widehat{T} \cap X = \{x^1, \dots, x^p\}$ and $\widehat{T} \cap Y = \{y^1, \dots, y^q\}$. Hence the convex hulls of $\widehat{T} \cap X$ and $\widehat{T} \cap Y$ intersect (the point z is contained in these convex hulls). Hence if $p+q \leq n+2$, then we have found our set T by defining $T = \widehat{T}$.

Now, let us consider the case $p+q > n+2$. Then the vectors $(\lambda_i x^i, \lambda_i, \lambda_i) \in \mathbb{R}^{n+2}$, $i = 1, \dots, p$, and $(\mu_j y^j, \mu_i, \mu_i) \in \mathbb{R}^{n+2}$, $j = 1, \dots, q$, are linearly dependent. (These are $p+q > n+2$ vectors in a $n+2$ dimensional space). Therefore, there exist vectors $\delta \in \mathbb{R}^p$ and $\theta \in \mathbb{R}^q$ with $(\delta, \theta) \neq 0$ and

$$\sum_{i=1}^{p} \delta_i \lambda_i x^i = \sum_{j=1}^{q} \theta_j \mu_j y^j,$$

$$\sum_{i=1}^{p} \delta_i \lambda_i = 0,$$

$$\sum_{j=1}^{q} \theta_j \mu_j = 0.$$

Since $(\delta, \theta) \neq 0$, we either have $\delta \neq 0$ or $\theta \neq 0$.

If $\delta \neq 0$, then at least one of the components of δ must be negative, that is, there exists some $i \in \{1, \dots, p\}$ with $\delta_i < 0$, as $\lambda_i > 0$ for all $i = 1, \dots, p$, $\delta \neq 0$, and $\sum_{i=1}^{p} \delta_i \lambda_i = 1$. Let us choose i^* with $\delta_{i^*} = \min_i \delta_i$ and define $\lambda_i^* = (1 - \delta_i/\delta_{i^*})\lambda_i$. We have $\delta_i/\delta_{i^*} \leq 1$ and hence $0 \leq \lambda_i^* \leq 1$ and $\lambda_{i^*}^* = 0$. We compute

$$\sum_{i=1}^{p} \lambda_i^* = \sum_{i=1}^{p} \lambda_i - (1/\delta_{i^*}) \sum_{i=1}^{p} \mu_i \lambda_i = 1 - 0,$$

$$\sum_{i \neq i^*} \lambda_i^* x^i = \sum_{i=1}^{p} \lambda_i^* x^i = \sum_{i=1}^{p} \lambda_i x^i - (1/\delta_{i^*}) \sum_{i=1}^{p} \mu_i \lambda_i x = x - 0 = x.$$

We obtain

$$\sum_{i \neq i^*} \lambda_i^* x^i = \sum_{i=1}^{p} \lambda_i x^i = z.$$

The set $S' = \widehat{T} \setminus \{x^{i^*}\}$ as $p+q-1$ elements. We have found a set S' with $p+q-1$ elements so that the convex hulls of $S' \cap X$ and $S' \cap Y$ have the point z in common. If $\theta \neq 0$, then we can apply the above computations to the y_j's and obtain a set T', by dropping one of the y_j's from $\widehat{T}$, with $p+q-1$ elements so that the convex hulls of $T' \cap X$ and $T' \cap Y$ have the point z in common. We can repeat the process of dropping points from $\widehat{T}$ until it only contains $n+2$ points.

Let us return to the proof of the Kirchberger's theorem. If the convex hulls of X and Y have a point in common, then the above considerations show that there exists a set $T \subset X \cup Y$ with at most $n+2$ elements such that $X \cap T$ and $Y \cap S$ intersect. If T has $n+2$ elements, we obtain a contradiction. If T has fewer elements than $n+2$, we can add as many elements from X and Y to T until we have found a set S with $n+2$ elements because the points $x^1, \ldots, x^k$ and $y^1, \ldots, y^m$ are distinct. We have $X \cap S \subset X \cap T$ and $Y \cap S \subset Y \cap T$. Since the convex hulls of $X \cap S$ and $Y \cap S$ have a point in common, the convex hulls of $X \cap T$ and $Y \cap T$ have a common point as well. We obtain a contradiction.

3 Homework 3

Exercise 2.1.
Prove Theorem 2.3.

Solution: Let $X \subset \mathbb{R}^n$ be nonempty and convex, and let $f\colon X \to \mathbb{R}$ be convex. Let $(x,t) \in \mathrm{epi}(f)$, $(y,s) \in \mathrm{epi}(f)$ and $\lambda \in [0,1]$. Since X and f are convex, we obtain

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \lambda t + (1-\lambda)s.$$

Hence

$$\lambda(x,t) + (1-\lambda)(y,s) = (\lambda x + (1-\lambda)y, \lambda t + (1-\lambda)s) \in \mathrm{epi}(f).$$

Now let X be nonempty and convex, and let $\mathrm{epi}(f)$ be convex. Let x, $y \in X$ and $\lambda \in [0,1]$. We have $(x, f(x)) \in \mathrm{epi}(f)$ and $(y, f(y)) \in \mathrm{epi}(f)$. Since $\mathrm{epi}(f)$ is convex, we have

$$\lambda(x, f(x)) + (1-\lambda)(y, f(y)) \in \mathrm{epi}(f).$$

Hence

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

Exercise 2.3.
Which of the following functions are convex on their indicated domains? No justifications are required.

1. $f(x) = 1$ on $\mathbb{R}$.
2. $f(x) = x$ on $\mathbb{R}$.
3. $f(x) = |x|$ on $\mathbb{R}$.
4. $f(x) = -|x|$ on $\mathbb{R}$.
5. $f(x) = -|x|$ on $\mathbb{R}_+ = \{x \in \mathbb{R}\colon x \geq 0\}$.
6. $f(x) = \exp(x)$ on $\mathbb{R}$.
7. $f(x) = \exp(x^2)$ on $\mathbb{R}$.
8. $f(x) = \exp(-x^2)$ on $\mathbb{R}$.
9. $f(x) = \exp(-x^2)$ on $\{x \in \mathbb{R}\colon x \geq 100\}$.
10. $f(x) = x\ln(x)$ on $\{x \in \mathbb{R}\colon x > 0\}$.

11. $f(x) = \sin(x)$ on $\mathbb{R}$.
12. $f(x) = -\ln(x)$ on $\{x \in \mathbb{R}\colon x > 0\}$.

Solution:

1. convex, as f is constant.
2. convex, as f is affine linear.
3. convex, as $f(x) = \max\{x, -x\}$ and $x \mapsto x$ and $x \mapsto -x$ are convex.
4. nonconvex and concave.
5. convex, as f is linear on the nonnegative x-axis.
6. convex, since $f''(x) = \exp(x) \geq 0$ for all $x \in \mathbb{R}$.
7. convex because $f''(x) = 2\exp(x^2)(2x^2+1) \geq 0$ for all $x \in \mathbb{R}$.
8. nonconvex, as $f''(0) = -2 < 0$.
9. convex, as $f''(x) = \exp(-x^2)(4x^2-2) \geq 0$ for $x \geq 100$.
10. convex, as $f''(x) = 1/x > 0$ for $x > 0$.
11. nonconvex, as $f''(x) = -\sin(x)$ and $f''(\pi/2) = -1$.
12. convex, as $f''(x) = 1/x^2$ for $x > 0$.

Exercise 2.4 (Maximum of a convex function over a polyhedron)**.**
Show that if f is a convex function on $\mathbb{R}^n$ and $X = \mathrm{Conv}(x^1, \ldots, x^m)$ with $x^i \in \mathbb{R}^n$, then

$$\sup_{x \in X} f(x) = \max_{i=1,\ldots,m} f(x^i).$$

Hint: Use Jensen's inequality.

Solution: Let $x \in X$. Then we can write x as a convex combination of $x^1, \ldots, x^m$, that is, $x = \sum_{i=1}^m \lambda_i x^i$ where $\lambda_i \geq 0$ and $\sum_{i=1}^m \lambda_i = 1$. Using Jensen's inequality, we obtain

$$f(x) = f\left(\sum_{i=1}^m \lambda_i x^i\right) \leq \sum_{i=1}^m \lambda_i f(x^i) \leq \max_{1 \leq i \leq m} f(x^i).$$

Hence

$$\sup_{x \in X} f(x) \leq \max_{i=1,\ldots,m} f(x^i)$$

Since $\{x^1, \ldots, x^m\}$ is a subset of its convex hull, we also have

$$\sup_{x \in X} f(x) \geq \sup_{x \in \{x^1,\ldots,x^m\}} f(x) = \max_{i=1,\ldots,m} f(x^i).$$

Exercise 2.6 (Products and ratios of convex functions)**.**
In general the product or ratio of two convex functions is not convex. However, there are some results that apply to functions on $\mathbb{R}$. Prove the following.

1. If f and g are convex, both nondecreasing, and positive functions on an interval, then fg is convex.

2. If f, g are concave, positive, with one nondecreasing and the other nonincreasing, then fg is concave.

3. If f is convex, nondecreasing, and positive, and g is concave, nonincreasing, and positive, then f/g is convex.

Solution:

1. Let $\lambda \in [0,1]$. Since f and g are positive and convex, we obtain

$$\begin{aligned} f(\lambda x + (1-\lambda)y)g(\lambda x + (1-\lambda)y) &\leq (\lambda f(x) + (1-\lambda)f(y))(\lambda g(x) + (1-\lambda)g(y)) \\ &= \lambda f(x)g(x) + (1-\lambda)f(y)g(y) \\ &\quad + \lambda(1-\lambda)(f(y) - f(x))(g(x) - g(y)). \end{aligned}$$

   The second term is less than or equal to zero, as f and g are nondecreasing. We obtain

$$f(\lambda x + (1-\lambda)y)g(\lambda x + (1-\lambda)y) \leq \lambda f(x)g(x) + (1-\lambda)f(y)g(y).$$

2. Reverse inequalities in the solution of part one.

3. It suffices to note that $1/g$ is convex, positive, and nondecreasing. The result follows from part one.

4 Homework 4

Exercise 2.13 (Strong convexity).
Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function and let $X \subset \mathrm{dom}(f)$ be a nonempty, convex set. Furthermore, let $\sigma > 0$ be a scalar. We say that f is *strongly convex over X with parameter (or coefficient) σ* if for all x, $y \in X$ and $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1 - \lambda)y) + (\sigma/2)\lambda(1 - \lambda)\|x - y\|_2^2 \leq \lambda f(x) + (1 - \lambda)f(y).$$

Note that the Euclidean norm is used in the above inequality.

Strongly convex functions are particularly "nice" functions when it comes to optimization. This problem establishes characterizations of strongly convex functions using first- and second-order derivative criteria, and implications of strong convexity. We use key properties of convex functions, such as the gradient inequality, to establish these characterizations. See Exercise 2.14 *for examples of strongly convex functions.*

1. Show that if f is strongly convex over X with coefficient σ, then f is strictly convex over X.
2. Show that f is strongly convex over X with coefficient σ if and only if the function $g(x) \coloneqq f(x) - (\sigma/2)\|x\|_2^2$ is convex over X.

   This characterization of strong convexity is **extremely** useful.
3. Suppose that $\mathrm{int}(X)$, the interior of X, is nonempty and that f is continuously differentiable on $\mathrm{int}(X)$. Show that the following statements are equivalent:
   a) f is strongly convex over X with parameter σ.
   b) We have
   $$f(y) \geq f(x) + \nabla f(x)^T(y - x) + (\sigma/2)\|x - y\|_2^2 \quad \text{for all} \quad x, y \in \mathrm{int}(X).$$
   c) We have
   $$(\nabla f(x) - \nabla f(y))^T(x - y) \geq \sigma\|x - y\|_2^2 \quad \text{for all} \quad x, y \in \mathrm{int}(X).$$

   *Hints: To establish "*a) *implies* b)*",* use *the second part and the gradient inequality. Exercise* 2.12 *may also be helpful.*
4. If in addition to the conditions in part three, f is twice continuously differentiable on $\mathrm{int}(X)$, then the conditions a)–c) are equivalent to $\nabla^2 f(x) - \sigma I$ is positive semidefinite for each $x \in \mathrm{int}(X)$.

Solution:

1. If $\lambda \in (0, 1)$ and $x \neq y$ with x, $y \in X$, then we have
$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &< f(\lambda x + (1 - \lambda)y) + (\sigma/2)\lambda(1 - \lambda)\|x - y\|_2^2 \\ &\leq \lambda f(x) + (1 - \lambda)f(y). \end{aligned}$$

   Hence f is strictly convex over X.

2. Let x, $y \in X$ and let $\lambda \in [0,1]$. In light of the definition of g, we only need to show that

$$-\|\lambda x + (1-\lambda)y\|_2^2 + \lambda\|x\|_2^2 + (1-\lambda)\|y\|_2^2 = \lambda(1-\lambda)\|x-y\|_2^2.$$

We compute

$$\begin{aligned}
-\|\lambda x + (1-\lambda)y\|_2^2 &+ \lambda\|x\|_2^2 + (1-\lambda)\|y\|_2^2 \\
&= \lambda(1-\lambda)\|x\|_2^2 - 2\lambda(1-\lambda)x^T y + \lambda(1-\lambda)\|y\|_2^2 \\
&= \lambda(1-\lambda)\|x-y\|_2^2.
\end{aligned}$$

3. "a) implies b)" Part two ensures that the function $g(x) = f(x) - (\sigma/2)\|x\|_2^2$ is convex. Let us fix $x \in \text{int}(X)$ and $y \in X$. The gradient inequality applied to g ensures

$$g(y) \geq g(x) + \nabla g(x)^T (y-x)$$

We have $\nabla g(x) = \nabla f(x) - \sigma x$. Hence

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) - \sigma x^T (y-x) - (\sigma/2)\|x\|_2^2 + (\sigma/2)\|y\|_2^2.$$

We further have

$$-\sigma x^T (y-x) - (\sigma/2)\|x\|_2^2 + (\sigma/2)\|y\|_2^2 = (\sigma/2)\|y-x\|_2^2.$$

To recognize this identity, we can use a second-order Taylor's expansion of $x \mapsto (1/2)\|y\|_2^2$ about x:

$$(1/2)\|y\|_2^2 = (1/2)\|x\|_2^2 + 2(1/2)x^T (y-x) + (1/2)\|x-y\|_2^2.$$

Alternatively, we can verify the identity using direct computations. Putting together the pieces, we have shown that "a) implies b)".

"b) implies c)" Fix x, $y \in \text{int}(X)$. Using part b), we obtain

$$\begin{aligned}
f(y) &\geq f(x) + \nabla f(x)^T (y-x) + (\sigma/2)\|x-y\|_2^2, \\
f(x) &\geq f(y) + \nabla f(y)^T (x-y) + (\sigma/2)\|x-y\|_2^2.
\end{aligned}$$

Adding the inequalities, we obtain

$$0 \geq (\nabla f(x) - \nabla f(y))^T (y-x) + 2(\sigma/2)\|x-y\|_2^2.$$

Rearranging terms yields c).

"c) implies a)" We apply Exercise 2.12 to the function $g(x) = f(x) - (\sigma/2)\|x\|_2^2$.

4. Applying Theorem 2.9 to the function $g(x) = f(x) - (\sigma/2)\|x\|_2^2$ ensures the assertions.

Exercise 2.14 (Examples of strongly convex functions).
This problem highlights an important strongly convex function. Moreover it establishes the fact that sums of convex and strongly convex functions are strongly convex.

1. Show that the function $(1/2)\|\cdot\|_2^2$ is strongly convex over $\mathbb{R}^n$ with parameter 1.
2. Show that if f is strongly convex over a nonempty convex set $X \subset \mathbb{R}^n$ with parameter $\sigma > 0$, and g is convex over X, then $f + g$ is strongly convex over X with parameter σ.

Solution:

1. The function $(1/2)\|\cdot\|_2^2 - (1/2)\|\cdot\|_2^2$ is zero and hence convex. Therefore, $(1/2)\|\cdot\|_2^2$ is strongly convex over $\mathbb{R}^n$ with parameter 1 (see Exercise 2.13).
2. The function $f(x) - (\sigma/2)\|x\|_2^2$ is convex over X (see Exercise 2.13) and hence $f(x) + g(x) - (\sigma/2)\|x\|_2^2$ is convex over X. Hence $f + g$ is strongly convex over X with parameter $\sigma > 0$ (see Exercise 2.13).

Exercise 2.15.
Provide counterexamples for the following statements.

1. If a function is strictly convex, then it is strongly convex.
2. If a function is strictly convex, then it has a minimizer.

Solution:

1. The function $f(x) = x^4$ is strictly convex, but not strongly convex.
2. The function $f(x) = \exp(x)$ is strictly convex, but it lacks a minimizer.

Exercise 2.16 (Minimization of strongly convex functions).
Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function that is continuous and strongly convex over a nonempty, closed, convex set X with parameter $\sigma > 0$.

Show that there exists a unique point $x^* \in X$ that minimizes f over X and that

$$f(x) \geq f(x^*) + (\sigma/2)\|x - x^*\|_2^2 \quad \text{for all} \quad x \in X.$$

Inequalities of this type are often referred to as quadratic growth conditions.

This problem demonstrates that every continuous, strongly convex function over a nonempty closed convex set has a unique minimizer and exhibits quadratic growth around it.

You may establish your own proof or solve the following subproblems.

1. Let us define $g(x) = f(x) - (\sigma/2)\|x\|_2^2$. Show that there exists $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that
$$g(x) \geq a^T x + b \quad \text{for all} \quad x \in X.$$
2. Let $x_0 \in X$. Deduce that the level set $\{x \in X \colon f(x) \leq f(x_0)\}$ is bounded.
3. Using the fact that a continuous function on a nonempty, closed, bounded set in $\mathbb{R}^n$ has a minimizer, show that f has a minimizer over X.
4. Complete the proof.

Solution:

1. The set X is nonempty and convex. Therefore, its relative interior is nonempty (see Theorem 1.16).

   The function g is convex and X is convex and has a point $\bar{x}$ in its relative interior. Therefore, there exists a subgradient a of g (see Proposition 2.27) and the subgradient inequality yields $g(x) \geq g(\bar{x}) + a^T(x - \bar{x})$ valid for all $x \in X$. Defining $b = g(\bar{x}) - a^T\bar{x}$ ensures the assertion.
2. We have $f(x) \geq a^T x + b + (\sigma/2)\|x\|_2^2$ for all $x \in X$. Suppose that the level set is unbounded. Then there exists a sequence $(x^k) \subset X$ with $f(x^k) \leq f(x_0)$ and $\|x^k\|_2 \to \infty$. Since $\sigma > 0$, we have
$$f(x^k) \geq a^T x^k + b + (\sigma/2)\|x^k\|_2^2 \to \infty \quad \text{as} \quad k \to \infty.$$
   Hence $f(x^k) \to \infty$ as $k \to \infty$. This contradicts $f(x^k) \leq f(x_0) < \infty$ for all $k \in \mathbb{N}$.
3. The level set is bounded as shown in part three and since f is continuous and X is closed, the level set is closed. Moreover, X is nonempty. Therefore, f has a minimizer over X.
4. Since f is strongly convex, it is strictly convex (see Exercise 2.13). Hence its minimizer $x^* \in X$ must be unique (see Proposition 2.17). Let $y \in X$. For all $\lambda \in (0, 1)$, the strong convexity of f ensures
$$f(\lambda x^* + (1 - \lambda)y) + (\sigma/2)\lambda(1 - \lambda)\|x^* - y\|_2^2 \leq \lambda f(x^*) + (1 - \lambda)f(y).$$
   Combined with $f(x^*) \leq f(\lambda x^* + (1 - \lambda)y)$ and $f(x^*) \in \mathbb{R}$, we obtain
$$(1 - \lambda)f(x^*) + (\sigma/2)\lambda(1 - \lambda)\|x^* - y\|_2^2 \leq (1 - \lambda)f(y).$$
   Dividing by $1 - \lambda$, we obtain
$$f(x^*) + (\sigma/2)\lambda\|x^* - y\|_2^2 \leq f(y).$$
   Taking limits as $\lambda \to 1$, we obtain the quadratic growth condition.

Exercise 3.17.
Compute the minimizer of the linear function

$$f(x) = c^T x$$

over the set

$$V_p = \left\{ x \in \mathbb{R}^n \colon \sum_{i=1}^{n} |x_i|^p \leq 1 \right\},$$

where $1 < p < \infty$.

Hint: Use the KKT conditions established in Theorem 3.7.

Solution: The optimization problem is convex and hence the KKT conditions sufficient optimality conditions. Since the feasible set satisfies Slater's condition, the KKT conditions are necessary and sufficient optimality conditions. See Theorem 3.7.

If $c = 0$, then any feasible point is optimal. Let now $c \neq 0$.

For $x \in \mathbb{R}^n$ and $\lambda \geq 0$, the KKT conditions are given by

$$c_i + \lambda^* p |x_i^*|^{p-1} \operatorname{sign}(x_i^*) = 0,\ i = 1, \ldots, n,\ \lambda^* \geq 0,\ \sum_{i=1}^{n} |x_i^*|^p \leq 1, \text{ and } \lambda^* = 0 \text{ if } \sum_{i=1}^{n} |x_i^*|^p < 1.$$

If $\sum_{i=1}^{n} |x_i^*|^p < 1$, then the KKT conditions yield $\lambda^* = 0$ and hence $\nabla f(x^*) = 0$ contradicting $c \neq 0$.

Now let $\sum_{i=1}^{n} |x_i^*|^p = 1$. Using $c_i + \lambda^* p |x_i^*|^{p-1} \operatorname{sign}(x_i^*) = 0$ for $i = 1, \ldots, n$, we obtain

$$|c_i| = \lambda^* p |x_i^*|^{p-1}. \tag{1}$$

Hence

$$|c_i|^{p/(p-1)} = (\lambda^*)^{p/(p-1)} p^{p/(p-1)} |x_i^*|^p.$$

Let us define

$$q := \frac{p}{p-1}.$$

Summing over $i = 1, \ldots, n$ and using $\sum_{i=1}^{n} |x_i^*|^p = 1$, we find that

$$\sum_{i=1}^{n} |c_i|^q = (\lambda^*)^q p^q.$$

Hence

$$\lambda^* = (1/p) \|c\|_q.$$

Combined with (1), we find that

$$|c_i| = \|c\|_q |x_i^*|^{p-1}.$$

Hence

$$|x_i^*| = \frac{|c_i|^{1/(p-1)}}{\|c\|_q^{1/(p-1)}}.$$

We also have

$$q - 1 = \frac{p - (p-1)}{p-1} = \frac{1}{p-1}.$$

We obtain

$$x_i^* = -\frac{|c_i|^{q-1}\operatorname{sign}(c_i)}{\|c\|_q^{q-1}}, \quad i = 1, \ldots, n.$$

Since x^* is a KKT point and the optimization problem is convex, x^* is a solution.

Exercise 3.18.
Let $a_1, \ldots, a_n > 0$ and let $\alpha, \beta > 0$. Solve the optimization problem

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^n \frac{a_i}{x_i^\alpha} \quad \text{s.t.} \quad x > 0, \quad \sum_{i=1}^n x_i^\beta \le 1.$$

Hint: Use the variable transformation $y_i = x_i^\beta$ to obtain a convex optimization problem, and use the KKT conditions to solve the optimization problem.

Solution: We transform the optimization problem into a convex problem using the variable transformation $y_i = x_i^\beta$. We obtain the problem

$$\min_{y \in \mathbb{R}^n} \sum_{i=1}^n a_i y_i^{-\alpha/\beta} \quad \text{s.t.} \quad y > 0, \quad \sum_{i=1}^n y_i \le 1.$$

We derive the KKT conditions for the new problem:

$$-(\alpha/\beta) a_i y_i^{-\alpha/\beta - 1} + \lambda^* = 0,\ i = 1, \ldots, n,\ y^* > 0,\ \lambda^* \ge 0,\ \sum_{i=1}^n y_i^* \le 1, \lambda^* = 0 \text{ if } \sum_{i=1}^n y_i^* < 1.$$

If $\sum_{i=1}^n y_i^* = 1$, then the KKT conditions yield

$$y_i^* = \frac{a_i^{\frac{\beta}{\alpha+\beta}}}{\sum_{j=1}^n a_j^{\frac{\beta}{\alpha+\beta}}}, \quad i = 1, \ldots, n.$$

Since the new optimization problem is convex and y^* is a KKT point, it must be a solution to the new problem.

Using our variable transformation, we obtain

$$x_i^* = \frac{a_i^{\frac{1}{\alpha+\beta}}}{\left(\sum_{j=1}^n a_j^{\frac{\beta}{\alpha+\beta}}\right)^{1/\beta}}, \quad i = 1, \ldots, n,$$

as a solution to the original problem.

5 Homework 5

Exercise 4.1.
Consider the inequality constrained problem depicted in Figure 1. Is x^* a KKT point?

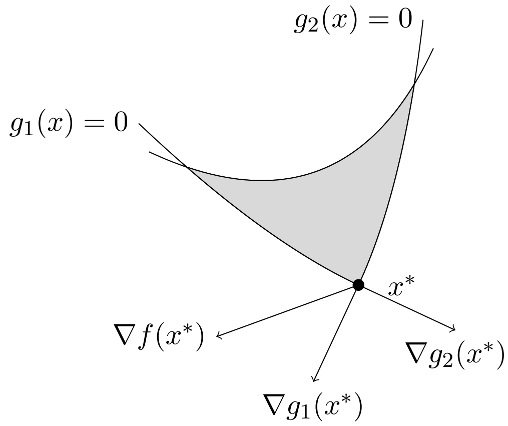

Figure 1: The gray region is the feasible set.

Solution: The vector $-\nabla f(x^*)$ is not a conic combination of the vectors $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$ and hence cannot be a KKT point.

Exercise 4.2.
We consider

$$\min_{x\in\mathbb{R}^2} f(x) \qquad \text{s.t.} \qquad h(x) = 0, \tag{P$_1$}$$

where $f : \mathbb{R}^2 \to \mathbb{R}$ and $h : \mathbb{R}^2 \to \mathbb{R}$ are defined by $f(x) := -x_1^4 x_2$ and $h(x) := -(x_1 + 2)^2 + x_2$.

Define $\bar{x} := (-2, 0) \in \mathbb{R}^2$ and $\bar{\mu} := 16$.

1. Show that $(\bar{x}, \bar{\mu})$ is a KKT tuple of problem (P$_1$).
2. Show that $\bar{x}$ is a regular point of (P$_1$)
3. Show that $T_+(h, \bar{x}, \bar{\mu}) = \{ d \in \mathbb{R}^2 : d_1 \in \mathbb{R},\, d_2 = 0 \}$.
4. Define $d := (1, 0)$. Show that $d^T \nabla_{xx} L(\bar{x}, \bar{\mu}) d = -32$, where L is the Lagrangian function corresponding to problem (P$_1$).
5. Show that $\bar{x}$ is not a local solution to problem (P$_1$).

Solution:

1. We have $h(\bar{x}) = -(-2+2)^2 + 0 = 0.$ and

$$\nabla f(x) = \begin{bmatrix} -4x_1^3 x_2 \\ -x_1^4 \end{bmatrix} \quad \text{and} \quad \nabla h(x) = \begin{bmatrix} -2(x_1+2) \\ 1 \end{bmatrix}.$$

Hence

$$\nabla f(\bar{x}) + \bar{\mu}\nabla h(\bar{x}) = \begin{bmatrix} 0 \\ -16 \end{bmatrix} + 16 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0.$$

2. We have $\nabla h(\bar{x}) \neq 0$. Hence $\bar{x}$ is a regular point.

3. We have $\nabla h(\bar{x}) = (0,1)$ and $T_+ = \{d \in \mathbb{R}^2 : \nabla h(\bar{x})^T d = 0\}$. Hence $T_+ = \{d \in \mathbb{R}^2 : d_1 \in \mathbb{R}, d_2 = 0\}$.

4. It suffices to compute $[\nabla_{xx} L(\bar{x}, \bar{\mu})]_{11}$. We have $[\nabla^2 f(x)]_{11} = -12x_1^2 x_2$ and $[\nabla^2 h(x)]_{11} = -2$. Hence

$$[\nabla^2 f(\bar{x})]_{11} = 0 \quad \text{and} \quad [\nabla^2 h(\bar{x})]_{11} = -2.$$

Since $L(x, \mu) = f(x) + \mu h(x)$, we have $d^T \nabla_{xx} L(\bar{x}, \bar{\mu}) d = -32 < 0$.

5. $(\bar{x}, \bar{\mu})$ is a KKT tuple and $\bar{x}$ is a regular point. If $\bar{x}$ would be a local solution, then the second-order necessary optimality conditions would hold. $(\bar{x}, \bar{\mu})$ is a KKT tuple. We have $d := (1, 0) \in T_+$. However, $d^T \nabla_{xx} L(\bar{x}, \bar{\mu}) d = -32 < 0$.

Alternative solution approach: Fix $\varepsilon > 0$ and define $x_\varepsilon = (\sqrt{\varepsilon} - 2, \varepsilon)^T$. We have $h(x_\varepsilon) = -(\sqrt{\varepsilon} - 2 + 2)^2 + \varepsilon = 0$ and $x_\varepsilon - \bar{x} = (\sqrt{\varepsilon}, \varepsilon)^T$. However, $f(\bar{x}) = 0 > f(x_\varepsilon)$.

Exercise 4.3.
We consider

$$\min_{x \in \mathbb{R}^3} f(x) \quad \text{s.t.} \quad g_1(x) \leq 0, \quad g_2(x) \leq 0, \quad h(x) = 0, \tag{P$_2$}$$

where $f(x) := x_1 + x_2 - 2x_3$, $g_1(x) := (1/2)x_1^2 - x_2$, $g_2(x) := \exp(x_1 - 1) - x_1$, and $h(x) := x_3^2 - x_1 + 3/4 = 0$.

We define $\bar{x} := (1, 1/2, 1/2) \in \mathbb{R}^3$, $\bar{\lambda} := (1, 0) \in \mathbb{R}^2$, and $\bar{\mu} := 2$.

1. Show that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple of (P$_2$).

2. Prove that $T_+(g, h, \bar{x}, \bar{\lambda}) = \{\, (t, t, t) : t \in \mathbb{R} \,\}$.

3. Show that

$$\nabla_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu}) = \begin{bmatrix} 1 & & \\ & 0 & \\ & & 4 \end{bmatrix}.$$

4. Show that $\bar{x}$ is a local solution to (P_2).

Solution:

1. We have $g_1(\bar{x}) = 1/2-1/2 = 0$, $g_2(\bar{x}) = \exp(1-1)-1 = 0$, $h(\bar{x}) = 1/4-1+3/4 = 0$, $\bar{\lambda} \geq 0$, $\bar{\lambda}_2 = 0$, $\bar{\lambda}_1 g_1(\bar{x}) = g_1(\bar{x}) = 0$, $\bar{\lambda}_2 g_2(\bar{x}) = \bar{\lambda}_2 \cdot 0 = 0$. Combined with $\bar{\lambda}_2 = 0$, we find that

$$\nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\mu} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + 1 \cdot \begin{bmatrix} \bar{x}_1 \\ -1 \\ 0 \end{bmatrix} + 2 \cdot \begin{bmatrix} -1 \\ 0 \\ 2\bar{x}_3 \end{bmatrix} = \begin{bmatrix} 1+1-2 \\ 0 \\ -2+4/2 \end{bmatrix} = 0.$$

2. Since $\nabla g_2(\bar{x}) = (\exp(0)-1, 0, 0) = (0,0,0)$, we have

$$\begin{aligned} T_+(g,h,\bar{x}) &= \{d \in \mathbb{R}^3 : \nabla g_1(\bar{x})^T d = \nabla h(\bar{x})^T d = 0\} = \{d \in \mathbb{R}^3 : d_1 - d_2 = 0, -d_1 + d_3 = 0\} \\ &= \{d \in \mathbb{R}^3 : d_1 = d_2 = d_3\}. \end{aligned}$$

3. The mappings f, g, and h are twice continuously differentiable, and $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple. Since $\bar{\lambda}_2 = 0$, we have

$$\begin{aligned} \nabla_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu}) &= \nabla^2 f(\bar{x}) + \bar{\lambda}_1 \nabla^2 g_1(\bar{x}) + \bar{\mu} \nabla^2 h(\bar{x}) \\ &= 0 + \begin{bmatrix} 1 & & \\ & 0 & \\ & & 0 \end{bmatrix} + 2 \begin{bmatrix} 0 & & \\ & 0 & \\ & & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & & \\ & 0 & \\ & & 4 \end{bmatrix}. \end{aligned}$$

Each $d \in T_+(h, \bar{x}) \setminus \{0\}$ can be written as $d = (t,t,t)$ with $t \in \mathbb{R} \setminus \{0\}$. For each $d = (t,t,t)$ with $t \in \mathbb{R} \setminus \{0\}$, we have

$$d^T \nabla_{xx}^2 L(\bar{x}, \bar{\mu}) d = 5t^2 > 0.$$

Hence the second-order sufficient optimality conditions impy that $\bar{x}$ is a strict local solution.

Exercise 4.4.
We consider the optimization problem

$$\min_{x \in \mathbb{R}^2} f(x) \quad \text{s.t.} \quad g_1(x) \leq 0, \quad g_2(x) \leq 0, \quad g_3(x) \leq 0, \tag{P$_3$}$$

where $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by $f(x) := x_1^3 x_2$ and $g : \mathbb{R}^2 \to \mathbb{R}$, $i = 1,2,3$ are given by

$$g_1(x) := x_1^2 + x_2^2 - 4, \quad g_2(x) := -(x_1+2)^2 + x_2, \quad g_3(x) := x_1 - \exp(-x_2).$$

We define $\bar{x} := (-2, 0) \in \mathbb{R}^2$ and $\bar{\lambda} := (0, 8, 0) \in \mathbb{R}^3$

1. Show that $\bar{x}$ is a KKT point of problem (P_3).
2. Show that $\bar{x}$ is a regular point of (P_3).
3. Show that $\bar{x}$ is not a local solution of problem (P_3).

Solution:

1. The KKT conditions for $(\bar{x}, \bar{\lambda})$ are given by
$$\begin{bmatrix} 0 \\ -8 \end{bmatrix} + \begin{bmatrix} -4 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \bar{\lambda} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \; g(\bar{x}) = (0, 0, -3) \leq 0, \bar{\lambda} \geq 0, \bar{\lambda}^T g(\bar{x}) = 0.$$
They are fulfilled only for $\bar{\lambda} = (0, 8, 0)$.

2. We have
$$\nabla g_{\mathcal{A}(\bar{x})}(\bar{x}) = \begin{bmatrix} -4 & 0 \\ 0 & 1 \end{bmatrix}.$$
Hence, $\bar{x}$ is a regular point.

3. If $\bar{x}$ would be local solution, then the second-order necessary conditions would be satisfied. The tuple $(\bar{x}, \bar{\lambda})$ is a KKT tuple. We have
$$\nabla_{xx} L(\bar{x}, \bar{\lambda}) = \begin{bmatrix} -16 & 12 \\ 12 & 0 \end{bmatrix}.$$
The critical cone is given by
$$T_+(g, \bar{x}, \bar{\lambda}) = \{\, d \in \mathbb{R}^2 : \, d_1 \geq 0, d_2 = 0 \,\}.$$
For $d := (1, 0)$, we have $d \in T_+(g, \bar{x}, \bar{\lambda})$ and $d^T \nabla_{xx} L(\bar{x}, \bar{\lambda}) d = -16 < 0$. Therefore, the second-order necessary optimality conditions are violated. Hence $\bar{x}$ cannot be a local solution of the problem.

Exercise 4.5 (Uniqueness of Lagrange multipliers under regularity)**.**
We consider
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_i(x) \leq 0, \; i = 1, \ldots, m \quad h_j(x) = 0, \; j = 1, \ldots, p, \tag{P_4}$$
where $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, and $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, p$, are continuously differentiable.

Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and $(\bar{x}, \widehat{\lambda}, \widehat{\mu})$ be KKT triples of (P_4). Suppose that $\bar{x}$ is *regular*. Show that $(\bar{\lambda}, \bar{\mu}) = (\widehat{\lambda}, \widehat{\mu})$.

Solution: The complementarity conditions yield $\bar{\lambda}_i = 0$ and $\widehat{\lambda}_i = 0$ for $i \notin \mathcal{A}(\bar{x})$. Subtracting the conditions $\nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0$ and $\nabla_x L(\bar{x}, \widehat{\lambda}, \widehat{\mu}) = 0$, we find that

$$\sum_{i \in \mathcal{A}(\bar{x})} \nabla g_i(\bar{x})(\bar{\lambda}_i - \widehat{\lambda}_i) + \sum_{j=1}^{p} \nabla h_j(\bar{x})(\bar{\mu}_j - \widehat{\mu}_j) = 0.$$

The regularity of $\bar{x}$ ensures that $\bar{\lambda}_i - \widehat{\lambda}_i = 0$, $i \in \mathcal{A}(\bar{x})$ and $\bar{\mu}_j - \widehat{\mu}_j = 0$, $j = 1, \ldots, p$ (because the gradients $\nabla\nabla g_i(\bar{x})$, $i \in \mathcal{A}(\bar{x})$ and $\nabla h_j(\bar{x}, j = 1, \ldots, p$ are linearly independent.) Putting together the pieces, we find that $(\bar{\lambda}, \bar{\mu}) = (\widehat{\lambda}, \widehat{\mu})$.

Exercise 4.6 (KKT conditions for the Celis–Dennis–Tapia problem)**.**
We consider

$$\min_{x \in \mathbb{R}^n} x^T H x + 2b^T x \quad \text{s.t.} \quad \|x\|_2^2 - \Delta^2 \leq 0, \quad \|A^T x + c\|_2^2 - \xi^2 \leq 0, \tag{P$_5$}$$

where $\Delta > 0$, $\xi \geq 0$, $H \in \mathbb{R}^{n \times n}$ is symmetric, $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}^m$.

Derive the KKT conditions for (P$_5$).

(You only need to state the KKT conditions for (P$_5$) and do not need to compute its KKT points. The KKT conditions for (P$_5$) generally lack closed-form solutions. As a consequence, the computation of KKT points requires numerical computations.)

Solution: Let $\bar{x} \in \mathbb{R}^n$. The KKT conditions for $\bar{x}$ are given by: there exists $\lambda \in \mathbb{R}^2$ such that

$$\begin{aligned}
2H\bar{x} + 2b + 2\lambda_1 \bar{x} + 2\lambda_2 A(A^T \bar{x} + c) &= 0, \\
\|\bar{x}\|_2^2 &\leq \Delta^2, \\
\|A^T \bar{x} + c\|_2^2 &\leq \xi^2, \\
\lambda &\geq 0, \\
\lambda_1(|\|\bar{x}\|_2^2 - \Delta^2) &= 0, \\
\lambda_2(\|A^T \bar{x} + c\|_2^2 - \xi^2) &= 0.
\end{aligned}$$

Exercise 4.7.
We consider the inequality constrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_i(x) \leq 0, \; i = 1, \ldots, m. \tag{P$_6$}$$

Let x^* be a local solution to (P$_6$) and let f and g_i, $i = 1, \ldots, m$, be continuously differentiable on $\mathbb{R}^n$. Suppose that the inequality constraints g_i, $i = 1, \ldots, m$, are concave.

1. Show that x^* is a local solution to the linearized optimization problem

$$\min_{x\in\mathbb{R}^n} f(x^*) + \nabla f(x^*)^T(x - x^*) \quad \text{s.t.} \quad g_i(x^*) + \nabla g_i(x^*)^T(x - x^*) \le 0,\ i = 1, \ldots, m.$$

2. Is x^* a KKT point of (P_6)?

Solution:

1. Let x be a feasible point for the linearized problem. Since the feasible set of the linearized problem is convex, the points $x_t = x^* + t(x - x^*)$ for $t \in [0, 1]$ are also feasible points for the linearized problem. The functions g_i are concave. So the gradient inequality and the feasibility of x_t imply

$$g_i(x_t) \le g_i(x^*) + \nabla g_i(x^*)^T(x_t - x^*) \le 0, \quad i = 1, \ldots, m.$$

   Therefore the points x_t are feasible for (P_6). By assumption x^* is a local solution to (P_6). Hence $f(x_t) \ge f(x^*)$ for all sufficiently small $t > 0$. We obtain for those t's,

$$0 \le \frac{f(x^* + t(x - x^*)) - f(x^*)}{t}.$$

   Taking limits as $t \to 0$, we obtain $\nabla f(x^*)^T(x - x^*) \ge 0$. Hence

$$f(x^*) + \nabla f(x^*)^T(x - x^*) \ge f(x^*) = \nabla f(x^*)^T(x^* - x^*) + f(x^*).$$

   Therefore x^* solves the linearized problem.

2. The point x^* is a KKT point of (P_6). This can be shown using the following approaches, for example.
   a) The point x^* solves the linearized problem. The linearized problem is an LP. The KKT conditions for this LP are the KKT conditions for the nonlinear problem (P_6).
   b) We can use the derivations in Section 2.7 applied to the linearized problem in order to obtain the KKT conditions for (P_6).
   c) We can use the derivations in Section 4.1.2.

6 Homework 6

Exercise 5.1 (Gradient descent with constant step size I)**.**
Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L \geq 0$. Let f^* be the optimal value of $\min_{x \in \mathbb{R}^n} f(x)$ and suppose that f^* is finite.

We consider the gradient descent method, Algorithm 5.1, with constant step size $\gamma_k = \gamma$.

Show that if $0 < \gamma < (2/L)$, then for all $K \in \mathbb{N}$, the sequence generated by the gradient descent satisfies

$$\min_{0 \leq k \leq K} \|\nabla f(x^k)\|_2^2 \leq \frac{1}{\gamma(1 - L\gamma/2)(K+1)} \cdot (f(x^0) - f^*).$$

You can establish your own proof or solve the following subproblems.

1. Show that for all $k \in \mathbb{N}$,
$$f(x^{k+1}) \leq f(x^k) - \gamma(1 - L\gamma/2)\|\nabla f(x^k)\|_2^2.$$
Hint: Use the descent lemma.
2. Show that
$$\gamma(1 - L\gamma/2) \sum_{k=0}^{K} \|\nabla f(x^k)\|_2^2 \leq f(x^0) - f(x^{K+1}).$$
3. Deduce the convergence rate.

Solution:

1. Choosing $y = x^k - \gamma \nabla f(x^k)$ and $x = x^k$ in the descent lemma, we obtain
$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) - \gamma\|\nabla f(x^k)\|_2^2 + \gamma^2(L/2)\|\nabla f(x^k)\|_2^2 \\ &= f(x^k) - \gamma(1 - L\gamma/2)\|\nabla f(x^k)\|_2^2. \end{aligned}$$
We obtain the assertion.
2. Using part one, we have
$$f(x^k) - f(x^{k+1}) \geq \gamma(1 - L\gamma/2)\|\nabla f(x^k)\|_2^2 \geq \gamma(1 - L\gamma/2)\|\nabla f(x^k)\|_2^2.$$
Summing over $k = 0, \ldots, K$ and using $\sum_{k=0}^{K} \left[f(x^k) - f(x^{k+1})\right] = f(x^0) - f(x^{K+1})$, we obtain the inequality.
3. We have $f(x^{K+1}) \geq f^*$. Hence $f(x^0) - f(x^K) \leq f(x^0) - f^*$. Moreover
$$\sum_{k=0}^{K} \|\nabla f(x^k)\|_2^2 \geq (K+1) \min_{0 \leq k \leq K} \|\nabla f(x^k)\|_2^2.$$
Putting together the pieces, we obtain the convergence rate.

Exercise 5.3 (Gradient descent applied strongly convex quadratic functions)**.**
Let $f : \mathbb{R}^n \to \mathbb{R}$ by defined by $f(x) := (1/2)x^T Ax - b^T x$, where $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite and $b \in \mathbb{R}^n$.

1. Show that $\nabla f(x) = Ax - b$.

2. Show that ∇f is Lipschitz continuous with Lipschitz constant $L = \lambda_{\max}(A)$ (maximum eigenvalue) and that f is strongly convex with parameter $\mu = \lambda_{\min}(A)$ (minimum eigenvalue).

3. Show that $x^* := A^{-1}b$ is the unique minimizer of f.

4. We consider the gradient descent with minimization rule applied to f. Let us define $g_k = Ax^k - b$. Show that if $g_k \neq 0$, then
$$\gamma_k = \frac{g_k^T g_k}{g_k^T A g_k}.$$

5. Consider the gradient method with constant step size $\gamma_k = \gamma > 0$ applied to f. Let L be the maximum eigenvalue of the matrix A. Show that the gradient method with constant step size γ converges to x^* for every starting point x^0 if and only if $\gamma \in (0, 2/L)$.

   Hint: Show that $x^{k+1} - x^* = (I - \gamma A)^{k+1}(x^0 - x^*)$.

Solution:

1. This identity follows from calculus rules, for example.

2. Using part one, we have
$$\begin{aligned}\|\nabla f(y) - \nabla f(x)\|_2^2 &= \|Ay - Ax\|_2^2 = \|A(y - x)\|_2^2 = (y - x)^T A^T A(y - x)\\ &\leq \lambda_{\max}(A^T A)\|y - x\|_2^2,\end{aligned}$$
   where $\lambda_{\max}(A^T A)$ is the maximum eigenvalue of $A^T A$. Since A is symmetric positive definite, we have $\sqrt{\lambda_{\max}(A^T A)} = \lambda_{\max}(A)$. Hence ∇f is Lipschitz continuous with Lipschitz constant $\lambda_{\max}(A)$.

   Alternatively, we compute
$$\|\nabla f(y) - \nabla f(x)\|_2 \|A(y - x)\|_2 \leq \|A\|_2\|y - x\|_2,$$
   where $\|A\|_2$ is the spectral norm of A. Since A is symmetric positive definite, $\|A\|_2$ equals its maximum eigenvalue.

3. f is a strongly convex function and hence has at most one minimizer. We have $\nabla f(x^*) = AA^{-1}b - b = 0$. Since f is convex, x^* is a minimizer of f.

4. We define $g = -\nabla f(x^k)$. Using a second-order Taylor's expansion of $\phi(\gamma) = f(x^k + \gamma g)$ about $\gamma = 0$, we find that

$$\begin{aligned} f(x^k + \gamma g) &= f(x^k) + \gamma \nabla f(x^k)^T g + (\gamma^2/2) g^T \nabla^2 f(x^k) g \\ &= f(x^k) - \gamma \|g\|_2^2 + (\gamma^2/2) g^T A g. \end{aligned}$$

   Hence $\phi'(\gamma) = 0$ if and only if $\|g\|_2^2 = \gamma g^T A g$. Since ϕ is strongly convex, $\gamma_k = \|g\|_2^2/(g^T A g)$ is the unique minimizer of ϕ over $\gamma \geq 0$. We obtain the assertion.

5. We have $Ax^* = b$ and hence $\nabla f(x^k) = A(x^k - x^*)$. Moreover,

$$\begin{aligned} x^{k+1} &= x^k - \gamma(Ax^k - b) = x^k - \gamma(Ax^k - Ax^*) = (I - \gamma A)x^k + \gamma A x^* \\ &= (I - \gamma A)x^k + (I + \gamma A - I)x^*. \end{aligned}$$

   Hence

$$x^{k+1} - x^* = (I - \gamma A)(x^k - x^*) = (I - \gamma A)^{k+1}(x^0 - x^*). \tag{2}$$

   We have $\lambda_{\min}(I - \gamma A) = 1 - \gamma \lambda_{\max}(A)$ and $\lambda_{\max}(I - \gamma A) = 1 - \gamma \lambda_{\min}(A)$.

   If $\gamma \in (0, 2/L)$ then the eigenvalues of $I - \gamma A$ are contained in $(-1, 1)$. Therefore, the matrix $(I - \gamma A)^{k+1}$ converges to the zero matrix 0 as $k \to \infty$.

   Now suppose that $x^{k+1} = x^k - \gamma \nabla f(x^k)$ converges to x^* for all $x^0 \in \mathbb{R}^n$ and let us show that $\gamma \in (0, 2/L)$. Let $y \in \mathbb{R}^n$ be an eigenvector of A with eigenvalue μ, that is, $Ay = \mu y$ and $y \neq 0$. We have for all $k \in \mathbb{N}$

$$(I - \gamma A)^{k+1} y = (1 - \gamma \mu)^{k+1} y.$$

   Let us choose $x^0 = x^* + y$. This choice yields $x^0 - x^* = y$. If $x^k \to x^*$, then (2) yields $(1 - \gamma\mu)^k y \to 0$. Since $y \neq 0$, there exists a component y_i of y with $|y_i| > 0$. Hence $(1 - \gamma\mu)^k \to 0$ as $k \to \infty$. Choosing once $\mu = \lambda_{\min}(A)$ and another time $\mu = \lambda_{\max}(A)$, we obtain $\gamma \in (0, 2/L)$.

Exercise 5.11 (Gradient descent with constant step size IV).
Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let its gradient ∇f be Lipschitz continuous with Lipschitz constant $L > 0$. Let f^* be the optimal value of $\min_{x \in \mathbb{R}^n} f(x)$ and suppose that f^* is finite.

Let $0 < \varrho < 2/(2 + L)$. We consider the gradient descent method, Algorithm 5.1, with constant step size γ_k fulfilling

$$\varrho \leq \gamma_k \leq \frac{2(1 - \varrho)}{L} \quad \text{for all} \quad k \in \mathbb{N}.$$

1. Show that for each $k \in \mathbb{N}$,

$$\varrho^2 \|\nabla f(x^k)\|_2^2 \leq f(x^k) - f(x^{k+1}).$$

   Hint: Use the descent lemma.

2. Show that for each $k \in \mathbb{N}$,

$$f(x^{k+1}) < f(x^k) \quad \text{provided that} \quad \nabla f(x^k) \neq 0.$$

3. Show that $(f(x^k))$ converges to some finite number $k \to \infty$.

4. Show that $\|\nabla f(x^k)\|_2 \to 0$ as $k \to \infty$.

5. Let $\bar{x}$ be an accumulation point of (x^k). Show that $\nabla f(\bar{x}) = 0$.

Solution:

1. Choosing $y = x^k - \gamma \nabla f(x^k)$ and $x = x^k$ in the descent lemma, we obtain

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) - \gamma_k \|\nabla f(x^k)\|_2^2 + \gamma_k^2 (L/2) \|\nabla f(x^k)\|_2^2 \\ &= f(x^k) - \gamma_k (1 - L\gamma_k/2) \|\nabla f(x^k)\|_2^2. \end{aligned}$$

Combined with $\gamma_k \geq \varrho$ and $(1 - L\gamma_k/2) \geq \varrho$, we obtain

$$\varrho^2 \|\nabla f(x^k)\|_2^2 \leq f(x^k) - f(x^{k+1}).$$

2. Using the first part, we obtain the assertion.

3. Since $(f(x^k))$ is nonincreasing and bounded from below, it converges to some limit $\geq f^*$.

4. We have $f(x^k) - f(x^{k+1}) \to 0$ as $k \to \infty$ (previous subproblem) and hence $\|\nabla f(x^k)\|_2 \to 0$ as $k \to \infty$ (see first subproblem).

5. Since ∇f is continuous, the previous subproblem yields $\nabla f(\bar{x}) = 0$.

Exercise 5.12 (Gradient descent with constant step size V)**.**
We define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) := \|x\|_2^2$. We consider the gradient descent method, Algorithm 5.1, with constant step size $\gamma_k = 1$ and initial point $x^0 := (1/2, 0)$.

1. Show that $\|x^k\|_2 = 1/2$ for all $k \in \mathbb{N}$.

2. Does the previous statement contradict the convergence statement in Exercise 5.11?

Solution:

1. We have $\nabla f(x) = 2x$. Hence $x^1 = x^0 - \nabla f(x^0) = -(1/2, 0)$. Moreover, $x^2 = x^1 - \nabla f(x^1) = (1/2, 0) = x^0$. Hence $\|x^k\|_2 = 1/2$ for all $k \in \mathbb{N}$.

2. The Lipschitz constant of ∇f is equal to 2. The step size conditions in Exercise 5.11 require
$$\gamma_k < 2/L = 1.$$
This is violated for $\gamma_k = 1$.

Exercise 5.15 (Can a gradient method with summable step sizes converge?)**.** Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, let $x^0 \in \mathbb{R}^n$, and let $S := \{x \in \mathbb{R}^n \colon f(x) \le f(x^0)\}$ be bounded. Denote by C the optimal value of $\max_{x\in S} \|\nabla f(x)\|_2$. (Why is C finite?) We consider the algorithm $x^{k+1} = x^k - \gamma_k \nabla f(x^k)$, where $\gamma_k \ge 0$. Suppose that $(x^k) \subset S$ and that $\Gamma := \sum_{k=0}^{\infty} \gamma_k < \infty$. Let $\bar{x}$ be a stationary point of f.

Note: (i) The results of this exercise have been used to construct Example 5.2. (ii) $(x^k) \subset S$ is true if $f(x^{k+1}) \le f(x^k)$ for all $k \in \mathbb{N} \cup \{0\}$, for example. (iii) $\sum_{k=0}^{\infty} \gamma_k < \infty$ is true if $\gamma_k := c/(1+k)^2$, where $c > 0$ is a constant, for example.

1. Show that for all $k \in \mathbb{N}$,
$$\|x^{k+1} - x^0\|_2 \le C\Gamma.$$
This means that $(x^k) \subset \mathrm{cl}(B_{C\Gamma}(x^0))$.
Hint: Use the fact that $x^{k+1} - x^0 = \sum_{i=0}^{k} (x^{i+1} - x^i)$.
Note: We have $\mathrm{cl}(B_{C\Gamma}(x^0)) = \{x \in \mathbb{R}^n \colon \|x - x^0\|_2 \le C\Gamma\}$ (the closed ball about x^0 with radius $C\Gamma$).
Solution: We have
$$x^{k+1} - x^0 = \sum_{i=0}^{k} x^{i+1} - x^i = -\sum_{i=0}^{k} \gamma_i \nabla f(x^i). \tag{3}$$
Combined with $\|\nabla f(x^i)\|_2 \le C$, we obtain the assertion.

2. Show that
$$\|x^{k+1} - \bar{x}\|_2 \ge \|\bar{x} - x^0\|_2 - C\Gamma.$$
Solution: We have
$$\|\bar{x} - x^0\|_2 \le \|x^{k+1} - x^0\|_2 + \|x^{k+1} - \bar{x}\|_2 \le C\Gamma + \|x^{k+1} - \bar{x}\|_2.$$

3. Deduce that if $\|\bar{x} - x^0\|_2 > C\Gamma$ then
$$\liminf_{k\to\infty} \|x^{k+1} - \bar{x}\|_2 > 0.$$
Solution: This is a consequence of the previous subproblem.