Applied Optimization: Formulation and Algorithms for Engineering Systems Slides

Ross Baldick

Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX 78712

Three introductory chapters

1
Introduction

1.1 Road map

- In this course, we are going to:
  - "formulate" various types of numerical problems, and
  - develop techniques for solving them.

Road map, continued

- We will use a number of case studies to:

  (i) illustrate the process of **formulating** a problem, that is, translating from an intuitive idea of the problem by writing it down mathematically,

  (ii) motivate and develop **algorithms** to solve problems, that is, descriptions of operations that can be implemented in software to take a problem specification and return a solution, and

  (iii) illustrate how to match the formulation of the problem to the capabilities of available algorithms, involving, in some cases, **transformation** of the problem from its initial formulation.

Road map, continued

- We will consider five general problem classes:
  (i) **linear systems of equations**,
  (ii) **non-linear systems of equations**,
  (iii) **unconstrained optimization**,
  (iv) **equality-constrained optimization**, and
  (v) **inequality-constrained optimization**.

Road map, continued

- Mostly consider problems that are defined in terms of **smooth** functions of **continuous** variables.
- We will emphasize issues that have proved pivotal in algorithm development and problem formulation:

  (i) **monotonicity**,
  (ii) **convexity**,
  (iii) **problem transformations**,
  (iv) **symmetry**, and
  (v) **sparsity**.

1.2 Goals

- At the end of the course, you should be able to:
  - (i) take a description of your problem,
  - (ii) translate it into a mathematical formulation,
  - (iii) evaluate if optimization techniques will be successful,
  - (iv) if the problem is tractable, solve small- to medium-scale versions of it using commercial code, and
  - (v) use the solution of the problem to calculate sensitivities to changes in problem specifications.

1.3 Pre-requisites

- The course assumes familiarity with MATLAB and MATLAB M-files and that you have access to the MATLAB Optimization Toolbox.
- In downloadable Appendix A there are notational conventions and a number of results that we will use in the course.

2

Problems, algorithms, and solutions

- In this chapter, we define the various types of problems that we will treat in the rest of the course.

Outline

- In Section 2.1, we define the **decision vector**.
- In Section 2.2 we define two problems involving **solution of simultaneous equations**.
- In Section 2.3 we describe three **optimization problems**.
- We define an **algorithm** in Section 2.4 in reference to two general schemata:
  - **direct** algorithms, which, in principle, obtain the exact solution to the problem in a finite number of operations, and
  - **iterative** algorithms, which generate a sequence of approximate solutions or “iterates” that, in principle, approach the exact solution to the problem.

2.1 Decision vector

- The problems will involve choices of a value of a **decision vector** from n-dimensional Euclidean space $\mathbb{R}^n$ or from some subset $\mathbb{S}$ of $\mathbb{R}^n$, where:
  - $\mathbb{R}$ is the set of real numbers, and
  - $\mathbb{R}^n$ is the set of n-tuples of real numbers.
- We will usually denote the decision vector by x.

2.2 Simultaneous equations

2.2.1 Definition

- Consider a vector function g that takes values of a decision vector in a **domain** $\mathbb{R}^n$ and returns values of the function that lie in a **range** $\mathbb{R}^m$.
- We write $g : \mathbb{R}^n \to \mathbb{R}^m$ to denote the domain and range of the function.
- Suppose we want to find a value $x^\star$ of the argument x that satisfies:

$$g(x) = \mathbf{0}. \tag{2.1}$$

- A value, $x^\star$, that satisfies (2.1) is called a solution of the **simultaneous equations** $g(x) = \mathbf{0}$.

Example

- Figure 2.1 shows the case of a function $g : \mathbb{R}^2 \to \mathbb{R}^2$.
- There are two sets illustrated by the solid curves.
- These two sets intersect at two points, $x^\star, x^{\star\star}$, illustrated as bullets $\bullet$.
- The points $x^\star$ and $x^{\star\star}$ are the two solutions of the simultaneous equations $g(x) = \mathbf{0}$, so that $\{x \in \mathbb{R}^n | g(x) = \mathbf{0}\} = \{x^\star, x^{\star\star}\}$.

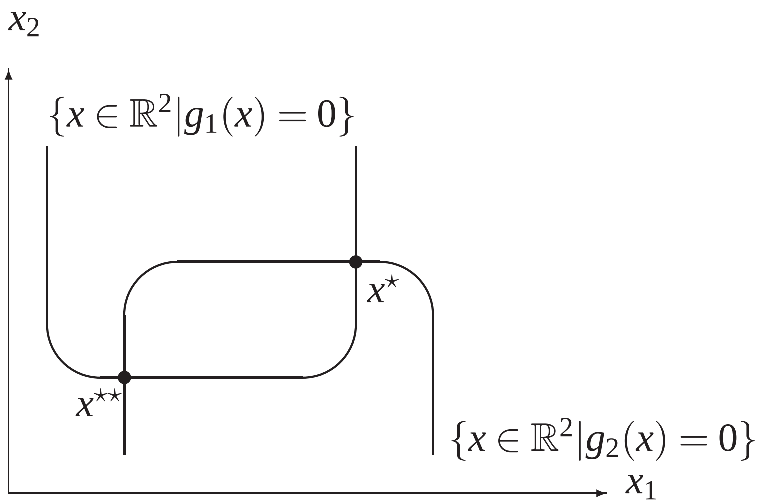


Fig. 2.1. Example of simultaneous equations and their solution.

Inconsistent and redundant equations

- If there are no solutions to the equations then $\{x \in \mathbb{R}^n | g(x) = \mathbf{0}\} = \emptyset$, where $\emptyset$ is the empty set, and we say that the equations are **inconsistent**.
- If some linear combination of the entries of g (with coefficients not all zero) yields a function that is identically zero then we say that the equations are **redundant**.
- For example, if two entries of g are the same then the equations are redundant.

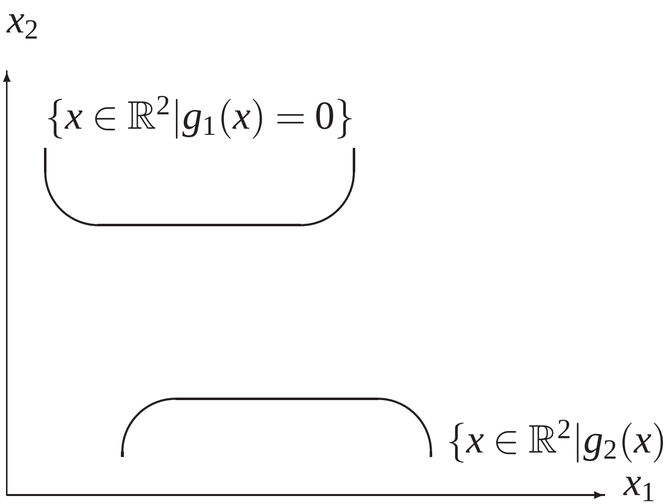


Fig. 2.2. Example of inconsistent simultaneous equations.

2.2.2 *Types of problems*

2.2.2.1 *Linear simultaneous equations*

- Suppose that $g : \mathbb{R}^n \to \mathbb{R}^m$ in (2.1) is **affine**, that is, of the form:

$$\forall x \in \mathbb{R}^n, g(x) = Ax - b.$$

- Then we have a set of **linear simultaneous equations:**

$$Ax - b = \mathbf{0}.$$

Examples

- For example, if:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \tag{2.2}$$

- then:

$$x^\star = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

- is a solution.

Examples, continued

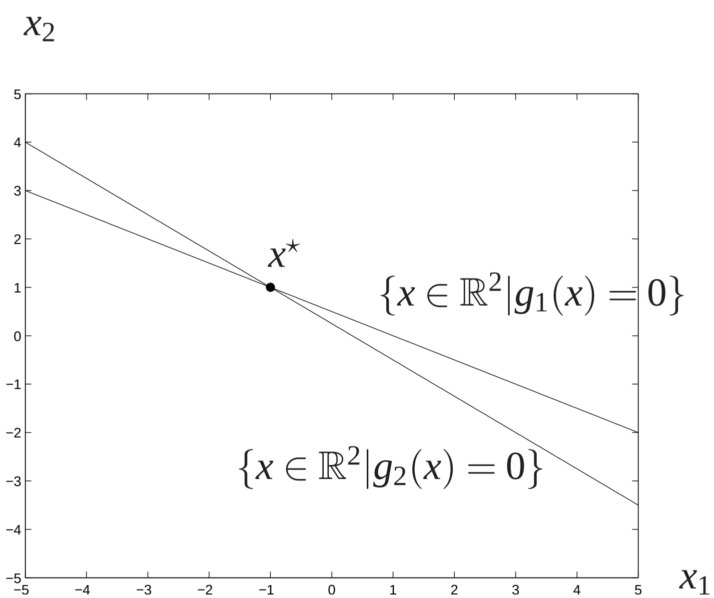


Fig. 2.3. Solution of linear simultaneous equations $g(x) = Ax - b = \mathbf{0}$ with A and b defined as in (2.2).

Examples, continued

- As another example, if:

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 7 & 6 & 5 \\ 8 & 9 & 11 \end{bmatrix}, b = \begin{bmatrix} 9 \\ 18 \\ 28 \end{bmatrix}, \tag{2.3}$$

- then:

$$x^\star = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- is a solution.

Number of solutions

- There may be several values that satisfy the equations.

Case studies

- Nodal analysis of a direct current linear circuit (in Section 4.1), and
- Control of a discrete-time linear system (in Section 4.2).

2.2.2.2 Non-linear simultaneous equations

Examples

- For example, suppose that the function $g : \mathbb{R} \to \mathbb{R}$ is defined by:

$$\forall x \in \mathbb{R}, g(x) = (x)^2 + 2x - 3. \tag{2.4}$$

- The "quadratic equation" shows that the two solutions are:

$$x^\star = -3, x^{\star\star} = 1.$$

Examples, continued

- As another example, let: $g : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by:

$$\forall x \in \mathbb{R}^2, g(x) = \begin{bmatrix} (x_1)^2 + (x_2)^2 + 2x_2 - 3 \\ x_1 - x_2 \end{bmatrix}. \qquad (2.5)$$

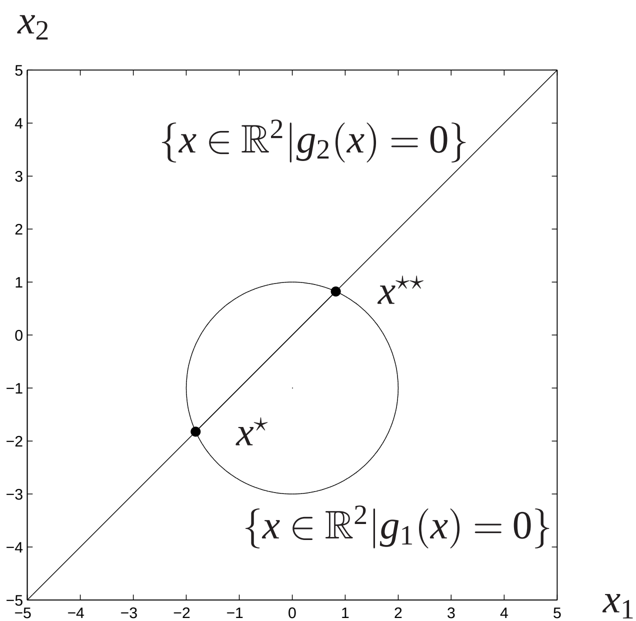


Fig. 2.4. Solution of non-linear simultaneous equations $g(x) = \mathbf{0}$ with g defined as in (2.5).

Examples, continued

- As a third example, let $g : \mathbb{R} \to \mathbb{R}$ be defined by:

$$\forall x \in \mathbb{R}, g(x) = (x-2)^3 + 1. \tag{2.6}$$

- By inspection, $x^\star = 1$ is the unique solution to $g(x) = 0$.

Algorithms and number of solutions

- Larger problems may also possess multiple solutions or no solutions under some circumstances.

Case studies

- Nodal analysis of a non-linear direct current electric circuit (in Section 6.1), and
- Analysis of an electric power system (in Section 6.2).

2.2.2.3 Eigenvalue problems

- Let $\mathbb{K}$ be the set of complex numbers.
- The n (not necessarily distinct) eigenvalues of a matrix $A \in \mathbb{R}^{n\times n}$ are given by the (possibly complex) solutions of the **characteristic equation** for A:

$$g(\lambda) = 0,$$

- where $g : \mathbb{K} \to \mathbb{K}$ is the **characteristic polynomial**, defined by:

$$\forall \lambda \in \mathbb{K}, g(\lambda) = \det(A - \lambda \mathbf{I}),$$

- The **eigenvectors** associated with an eigenvalue λ are the solutions of:

$$(A - \lambda \mathbf{I})x = \mathbf{0}.$$

Example

$$A = \begin{bmatrix} 2 & 1 \\ -5 & -4 \end{bmatrix},$$

$$\begin{aligned}
\forall \lambda \in \mathbb{K}, g(\lambda) &= \det(A - \lambda \mathbf{I}), \\
&= \det \begin{bmatrix} 2-\lambda & 1 \\ -5 & -4-\lambda \end{bmatrix}, \\
&= (2-\lambda)(-4-\lambda) - (1)(-5), \\
&= (\lambda)^2 + 2\lambda - 3.
\end{aligned}$$

- From the previous example, we already know that the two solutions to $g(\lambda) = 0$ are:

$$\lambda^\star = -3, \lambda^{\star\star} = 1,$$

- so these are the eigenvalues of A.

Example, continued

- The eigenvectors associated with $\lambda^\star = -3$ are the vectors in the set:

$$\{x \in \mathbb{R}^2 | (A + 3\mathbf{I})x = \mathbf{0}\}.$$

- The eigenvectors associated with $\lambda^{\star\star} = 1$ are the vectors in the set:

$$\{x \in \mathbb{R}^2 | (A - \mathbf{I})x = \mathbf{0}\}.$$

Discussion

- There are special iterative algorithms for eigenvalue problems that are somewhat different in flavor to the algorithms we will describe for solving general linear and non-linear equations.
- We will not discuss general algorithms for eigenvalue calculation.

2.3 Optimization

2.3.1 Definitions

2.3.1.1 Objective

- Consider a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ that denominates the “cost” or lack of desirability of solutions for a particular model or system.
- That is, $f(x)$ is the cost of using x as the solution.
- The function is called an **objective function**.

Example

- An example of a **quadratic** function $f : \mathbb{R}^2 \to \mathbb{R}$ is given by:

$$\forall x \in \mathbb{R}^2, f(x) = (x_1)^2 + (x_2)^2 + 2x_2 - 3. \tag{2.7}$$

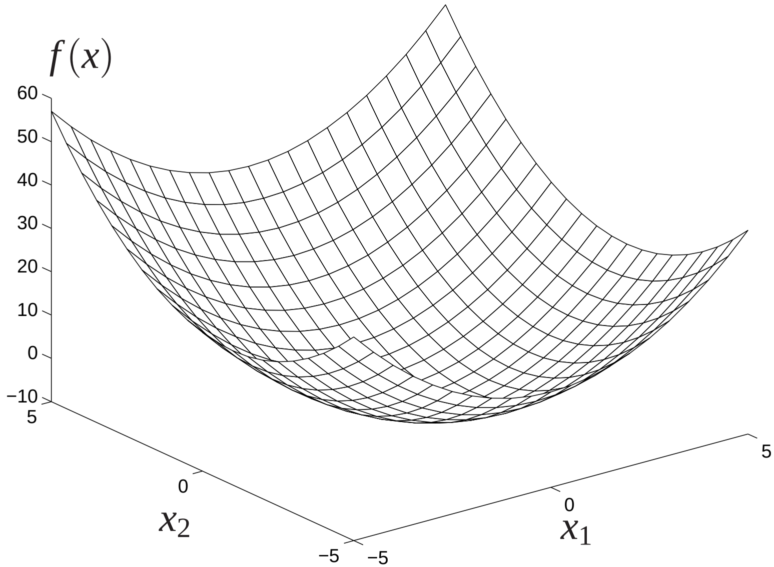


Fig. 2.5. Graph of the example objective function defined in (2.7).

Discussion

- We can categorize objectives according to the highest power of any entry in the argument.
- We will categorize objectives in a different way in Section 2.6.3.4 once we have discussed optimization in more detail.

2.3.1.2 Feasible set

- Our problem might involve restrictions on the choices of values of x.
- We can imagine a **feasible set** $\mathbb{S} \subseteq \mathbb{R}^n$ from which we must select a solution.

2.3.1.3 Problem

- A **minimization problem** means to find the minimum value of $f(x)$ over choices of x that lie in the feasible set $\mathbb{S}$.

Definition 2.1 Let $\mathbb{S} \subseteq \mathbb{R}^n$, $f : \mathbb{S} \to \mathbb{R}$, and $f^\star \in \mathbb{R}$. Then by:

$$f^\star = \min_{x \in \mathbb{S}} f(x), \tag{2.8}$$

we mean that:

$$\exists x^\star \in \mathbb{S} \text{ such that: } (f^\star = f(x^\star)) \text{ and } ((x \in \mathbb{S}) \Rightarrow (f(x^\star) \leq f(x))). \tag{2.9}$$

□

2.3.1.4 Set of minimizers

- The set of *all* the minimizers of $\min_{x\in\mathbb{S}} f(x)$ is denoted by:

$$\operatorname*{argmin}_{x\in\mathbb{S}} f(x).$$

- If the problem has no minimum (and, therefore, no minimizers) then we define:

$$\operatorname*{argmin}_{x\in\mathbb{S}} f(x) = \emptyset.$$

- To emphasize the role of $\mathbb{S}$, we also use the following notations:

$$\min_{x\in\mathbb{R}^n}\{f(x)|x\in\mathbb{S}\} \text{ and } \operatorname*{argmin}_{x\in\mathbb{R}^n}\{f(x)|x\in\mathbb{S}\}.$$

- We will often use a more explicit notation if $\mathbb{S}$ is defined as the set of points satisfying a criterion.
- For example, if $f:\mathbb{R}^n\to\mathbb{R}$, $g:\mathbb{R}^n\to\mathbb{R}^m$, $h:\mathbb{R}^n\to\mathbb{R}^r$, and $\mathbb{S}=\{x\in\mathbb{R}^n|g(x)=\mathbf{0},h(x)\leq\mathbf{0}\}$ then we will write $\min_{x\in\mathbb{R}^n}\{f(x)|g(x)=\mathbf{0},h(x)\leq\mathbf{0}\}$ for $\min_{x\in\mathbb{S}} f(x)$.

Multiple minimizers

- For example, consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}, f(x) = |x+1|^{1.5}(x-1)^3 + 1.$$

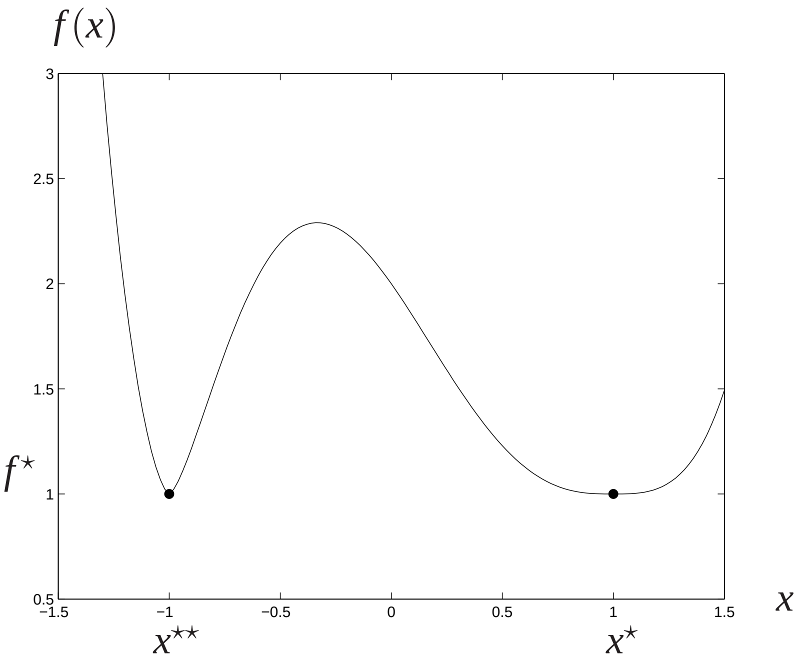


Fig. 2.6. Function having multiple unconstrained minimizers indicated by the bullets •.

2.3.1.5 Lower bound

Definition 2.2 Let $\mathbb{S} \subseteq \mathbb{R}^n$, $f : \mathbb{S} \to \mathbb{R}$, and $\underline{f} \in \mathbb{R}$. If $\underline{f}$ satisfies:

$$\forall x \in \mathbb{S}, \underline{f} \leq f(x),$$

then we say that $\underline{f}$ is a **lower bound** for the problem $\min_{x \in \mathbb{S}} f(x)$ or that the problem $\min_{x \in \mathbb{S}} f(x)$ is **bounded below** by $\underline{f}$. If $\mathbb{S} \neq \emptyset$ but no such $\underline{f}$ exists, then we say that the problem $\min_{x \in \mathbb{S}} f(x)$ is **unbounded below** (or unbounded if the "below" is clear from context.) $\square$

- Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined in (2.7), which we repeat here:

$$\forall x \in \mathbb{R}^2, f(x) = (x_1)^2 + (x_2)^2 + 2x_2 - 3.$$

- This function is illustrated in Figure 2.5.
- For the feasible set $\mathbb{S} = \mathbb{R}^2$, the value $\underline{f} = -10$ is a lower bound for the problem $\min_{x \in \mathbb{S}} f(x)$, as shown in Figure 2.5.

2.3.1.6 Level and contour sets

Definition 2.3 Let $\mathbb{S} \subseteq \mathbb{R}^n$, $f : \mathbb{S} \to \mathbb{R}$, and $\tilde{f} \in \mathbb{R}$. Then the **level set** at value $\tilde{f}$ of the function f is the set:

$$\mathbb{L}_f(\tilde{f}) = \{x \in \mathbb{S} | f(x) \leq \tilde{f}\}.$$

The **contour set** at value $\tilde{f}$ of the function f is the set:

$$\mathbb{C}_f(\tilde{f}) = \{x \in \mathbb{S} | f(x) = \tilde{f}\}.$$

For each possible function f, we can think of $\mathbb{L}_f$ and $\mathbb{C}_f$ themselves as *set-valued functions* from $\mathbb{R}$ to $(2)^{(\mathbb{R}^n)}$, where $(2)^{(\mathbb{R}^n)}$ denotes the **set of all subsets of** $\mathbb{R}^n$, sometimes called the **power set** of $\mathbb{R}^n$. $\square$

Example

- Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}^2, f(x) = (x_1 - 1)^2 + (x_2 - 3)^2. \tag{2.10}$$

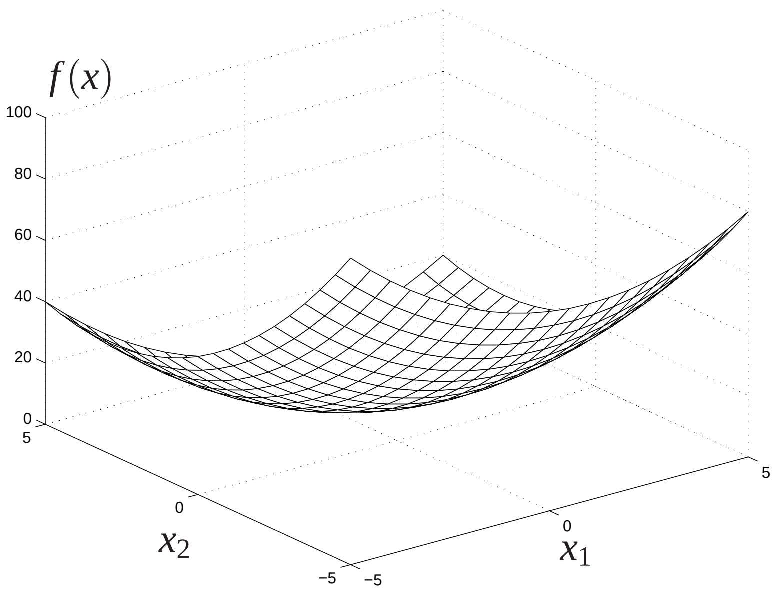


Fig. 2.7. Graph of function defined in (2.10).

Contour set for example

- The contour sets $\mathbb{C}_f(\tilde{f})$ can be shown in a two-dimensional representation.

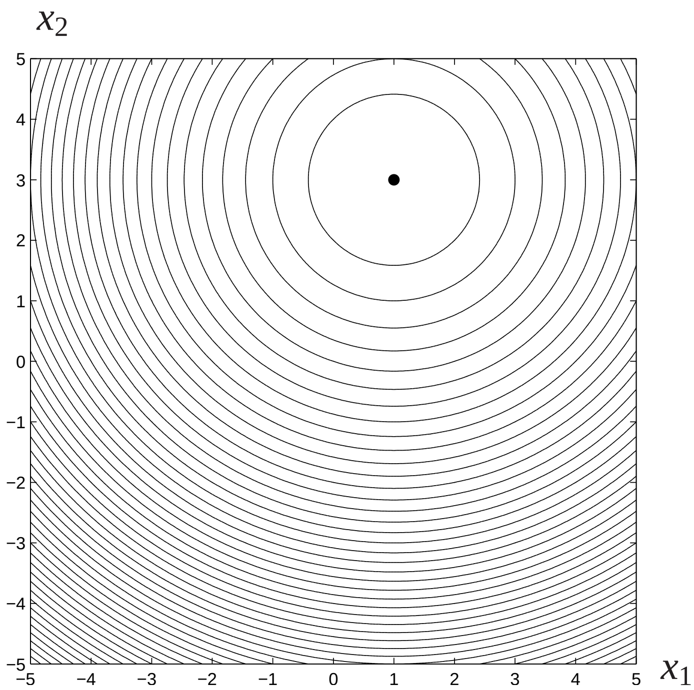


Fig. 2.8. Contour sets $\mathbb{C}_f(\tilde{f})$ of the function defined in (2.10) for values $\tilde{f} = 0, 2, 4, 6, \ldots$. The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$, which is illustrated with a $\bullet$ and is the contour of height 0.

2.3.2 *Types of problems*

- The three general forms of $\mathbb{S}$ that we will consider are:
  - **unconstrained** optimization,
  - **equality-constrained** optimization, and
  - **inequality-constrained** optimization.

2.3.2.1 Unconstrained optimization

- If $\mathbb{S} = \mathbb{R}^n$ then the problem is said to be **unconstrained**.

Example

- For example, consider the objective $f : \mathbb{R}^2 \to \mathbb{R}$ defined in (2.10):

$$\forall x \in \mathbb{R}^2, f(x) = (x_1 - 1)^2 + (x_2 - 3)^2.$$

- From Figure 2.8, which shows the contour sets of f, we can see that:

$$\begin{aligned} \min_{x \in \mathbb{R}^2} f(x) &= f^\star = 0, \\ \operatorname*{argmin}_{x \in \mathbb{R}^2} f(x) &= \left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\}, \end{aligned}$$

- so that there is a minimum $f^\star = 0$ and a unique minimizer $x^\star = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ of this problem.

Another example

- Consider a linear system $Ax - b = \mathbf{0}$ that does not have a solution.
- We may try to seek a value of the decision vector that "most nearly" satisfies $Ax = b$ in the sense of minimizing a criterion.
- A natural criterion is to consider a **norm** $\|\bullet\|$ and then seek x that minimizes $\|Ax - b\|$:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\| . \tag{2.11}$$

Case studies

- Multi-variate linear regression (in Section 9.1), and
- Power system state estimation (in Section 9.2).

2.3.2.2 Equality-constrained optimization

- If $g : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbb{S} = \{x \in \mathbb{R}^n | g(x) = \mathbf{0}\}$ then the problem is said to be **equality-constrained.**

Sub-types of equality-constrained optimization problems

Linearly constrained

- If g is affine then the problem is called **linearly constrained**.

Example

$$\begin{aligned} \forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\ \forall x \in \mathbb{R}^2, g(x) &= x_1 - x_2, \qquad (2.12) \\ \min_{x \in \mathbb{R}^2} \{f(x) | g(x) = 0\} &= \min_{x \in \mathbb{R}^2} \{f(x) | x_1 - x_2 = 0\}. \qquad (2.13) \end{aligned}$$

Example, continued

- The unique minimizer of Problem (2.13) is $x^\star = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

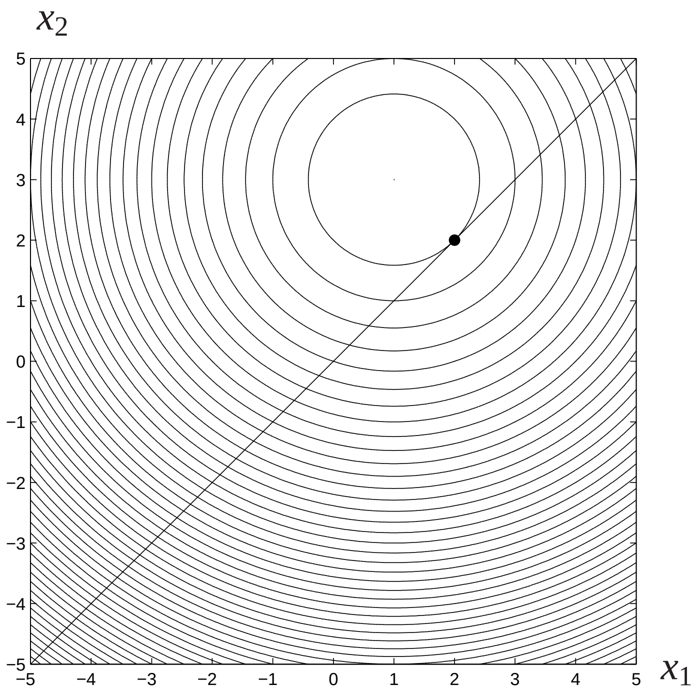


Fig. 2.9. Contour sets $\mathbb{C}_f(\tilde{f})$ of function repeated from Figure 2.8 with feasible set from Problem (2.13) superimposed. The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. The minimizer $x^\star = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ is illustrated with a $\bullet$.

Non-linearly constrained

- If there is no restriction on g then the problem is called **non-linearly constrained**.

Example

- For example, consider the same objective as previously, $f : \mathbb{R}^2 \to \mathbb{R}$ defined in (2.10):

$$\forall x \in \mathbb{R}^2, f(x) = (x_1 - 1)^2 + (x_2 - 3)^2.$$

- However, let $g : \mathbb{R}^2 \to \mathbb{R}$ be defined by:

$$\forall x \in \mathbb{R}^2, g(x) = (x_1)^2 + (x_2)^2 + 2x_2 - 3.$$

- Consider the equality-constrained problem:

$$\min_{x \in \mathbb{R}^2} \{f(x) | g(x) = 0\}. \tag{2.14}$$

- The unique minimizer of Problem (2.14) is $x^\star \approx \begin{bmatrix} 0.5 \\ 0.9 \end{bmatrix}$.

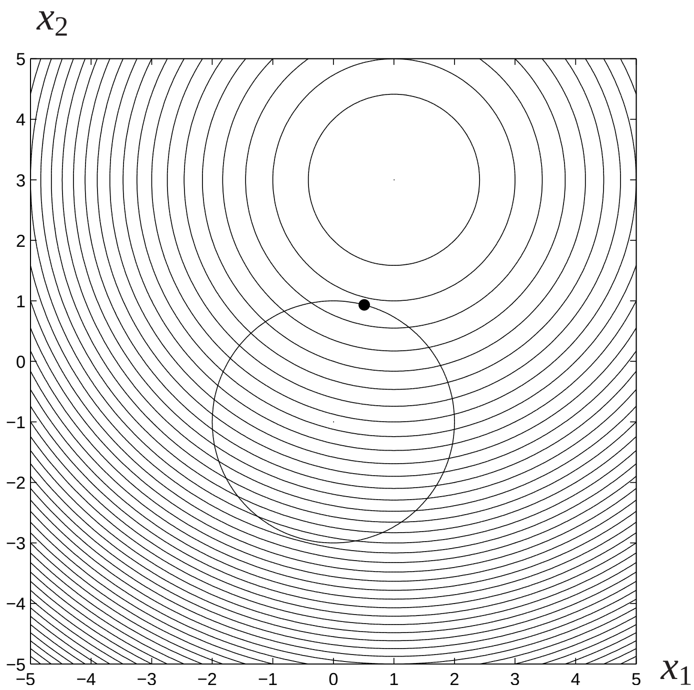


Fig. 2.10. Contour sets $\mathbb{C}_f(\tilde{f})$ of function repeated from Figure 2.8 with feasible set from Problem (2.14) superimposed. The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. The minimizer $x^\star$ is illustrated as a $\bullet$.

Case studies

- Least-cost production of a group of manufacturing facilities that must collectively meet a demand constraint (in Section 12.1), and
- Power system state estimation with zero injection buses (in Section 12.2).

2.3.2.3 *Inequality-constrained optimization*

- If $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^r$, and $\mathbb{S} = \{x \in \mathbb{R}^n | g(x) = \mathbf{0}, h(x) \leq \mathbf{0}\}$ then the problem is said to be **inequality-constrained.**

Sub-types of inequality-constrained optimization problems

Non-negatively constrained

- If h is of the form:

$$\forall x, h(x) = -x,$$

- so that the constraints are of the form $x \geq \mathbf{0}$ then the problem is **non-negatively constrained**.

Linear inequality constraints

- If h is affine then the problem is **linear inequality-constrained**.

Linear program

- If the objective is linear and g and h are affine then the problem is called a **linear program** or a **linear optimization problem**.

Example

$$
\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= x_1 - x_2, \\
\forall x \in \mathbb{R}^2, g(x) &= x_1 + x_2 - 1, \\
\forall x \in \mathbb{R}^2, h(x) &= \begin{bmatrix} -x_1 \\ -x_2 \end{bmatrix},
\end{aligned}
$$

$$
\min_{x \in \mathbb{R}^2} \{ f(x) | g(x) = 0, h(x) \leq \mathbf{0} \} = \min_{x \in \mathbb{R}^2} \{ x_1 - x_2 | x_1 + x_2 - 1 = 0, x_1 \geq 0, x_2 \geq 0 \}. \tag{2.15}
$$

Example, continued

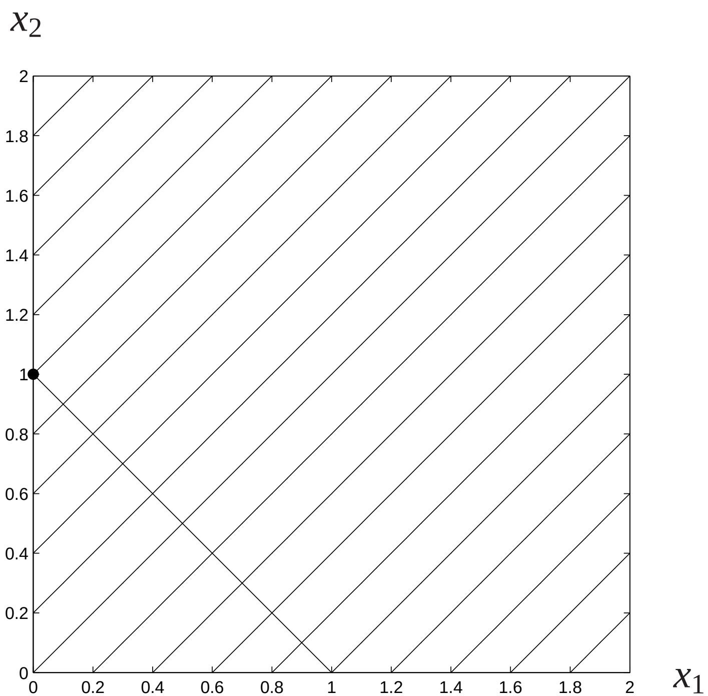


Fig. 2.11. Contour sets $\mathbb{C}_f(\tilde{f})$ of objective function and feasible set for Problem (2.15). The contour sets are the parallel lines. The feasible set is shown as the line joining the two points $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. The heights of the contours decrease to the left and up. The minimizer $x^\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is illustrated as a $\bullet$.

- We often emphasize the linear and affine functions by writing:

$$\min_{x\in\mathbb{R}^2}\{c^\dagger x | Ax = b, Cx \leq d\},$$

- where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{r\times n}$, and $d \in \mathbb{R}^r$.
- For Problem (2.15), the appropriate vectors and matrices are:

$$c = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, A = \begin{bmatrix} 1 & 1 \end{bmatrix}, b = [1], C = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

- We can write this non-negatively constrained problem even more concisely as:

$$\min_{x\in\mathbb{R}^2}\{c^\dagger x | Ax = b, x \geq \mathbf{0}\}. \tag{2.16}$$

Linear program, continued

- There is a rich body of literature on linear programming and there are special purpose algorithms to solve linear programming problems.
- The best known are:
  - the **simplex algorithm** (and variants), and
  - **interior point algorithms**.

Standard format

- If g is affine and the inequality constraints are non-negativity constraints then the problem is said to be in the **standard format**.
- Problem (2.16) is a linear program in standard format.

Quadratic program

- If f is quadratic and g and h are affine then the problem is called a **quadratic program** or a **quadratic optimization problem**.

Example

$$\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\
\forall x \in \mathbb{R}^2, g(x) &= x_1 - x_2, \\
\forall x \in \mathbb{R}^2, h(x) &= 3 - x_2.
\end{aligned} \tag{2.17}$$

Example, continued

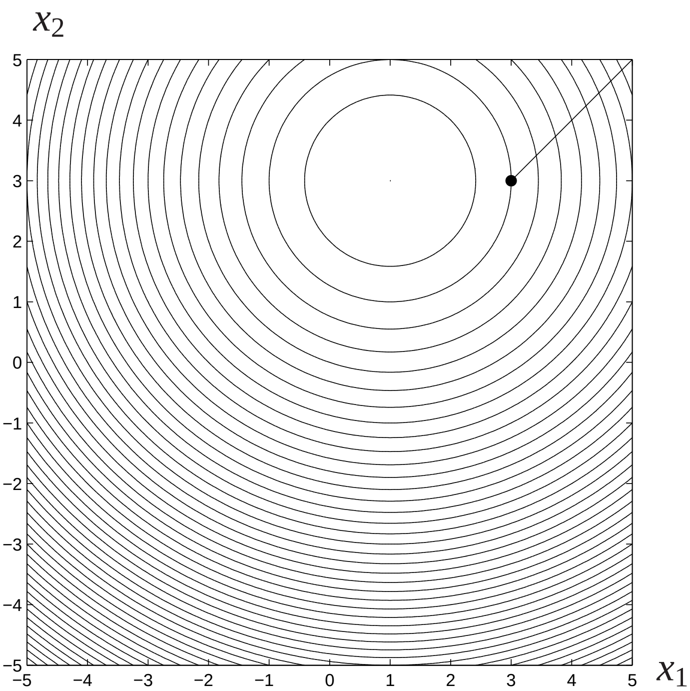


Fig. 2.12. Contour sets $\mathbb{C}_f(\tilde{f})$ of objective function and feasible set for Problem (2.18). The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. The feasible set is the "half-line" starting at the point $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$. The minimizer $x^\star = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ is illustrated with a $\bullet$.

Example, continued

$$\min_{x\in\mathbb{R}^2}\{f(x)|g(x)=0,h(x)\leq 0\} = 4, \tag{2.18}$$

$$\operatorname{argmin}_{x\in\mathbb{R}^2}\{f(x)|g(x)=0,h(x)\leq 0\} = \left\{\begin{bmatrix}3\\3\end{bmatrix}\right\} = \{x^\star\}.$$

Quadratic program, continued

- We can emphasize the quadratic and linear functions by writing:

$$\min_{x\in\mathbb{R}^2}\left\{\frac{1}{2}x^\dagger Qx+c^\dagger x|Ax=b,Cx\leq d\right\},$$

- where we have omitted the constant term.
- For Problem (2.18), the appropriate vectors and matrices are:

$$Q=\begin{bmatrix}2 & 0\\0 & 2\end{bmatrix},c=\begin{bmatrix}-2\\-6\end{bmatrix},A=\begin{bmatrix}1 & -1\end{bmatrix},b=[0],C=\begin{bmatrix}0 & -1\end{bmatrix},d=[-3].$$

Non-linear program

- If there are no restrictions on f, g, and h, then the problem is called a **non-linear program** or a **non-linear optimization problem**.

Example

$$\min_{x \in \mathbb{R}^3} \{ f(x) | g(x) = 0, h(x) \leq 0 \}, \tag{2.19}$$

- where $f : \mathbb{R}^3 \to \mathbb{R}$, $g : \mathbb{R}^3 \to \mathbb{R}^2$, and $h : \mathbb{R}^3 \to \mathbb{R}$ are defined by:

$$\begin{aligned} \forall x \in \mathbb{R}^3, f(x) &= (x_1)^2 + 2(x_2)^2, \\ \forall x \in \mathbb{R}^3, g(x) &= \begin{bmatrix} 2 - x_2 - \sin(x_3) \\ -x_1 + \sin(x_3) \end{bmatrix}, \\ \forall x \in \mathbb{R}^3, h(x) &= \sin(x_3) - 0.5. \end{aligned}$$

Convexity

- We will see in Section 2.6.3 that we can also classify problems on the basis of the notion of **convexity**.

Satisfaction of constraints

Definition 2.4 Let $h : \mathbb{R}^n \to \mathbb{R}^r$. An inequality constraint $h_\ell(x) \leq 0$ is called a **binding constraint** or an **active constraint** at $x^\star$ if $h_\ell(x^\star) = 0$. It is called **non-binding** or **inactive** at $x^\star$ if $h_\ell(x^\star) < 0$. The set:

$$\mathbb{A}(x^\star) = \{\ell \in \{1, \ldots, r\} \,|h_\ell(x^\star) = 0\}$$

is called the **set of active constraints** or the **active set** for $h(x) \leq \mathbf{0}$ at $x^\star$. $\square$

Definition 2.5 Let $h : \mathbb{R}^n \to \mathbb{R}^r$. The point $x^\star$ is called **strictly feasible** for the inequality constraint $h_\ell(x) \leq 0$ if $h_\ell(x^\star) < 0$. The point $x^\star$ is called **strictly feasible** for the inequality constraints $h(x) \leq \mathbf{0}$ if $h(x^\star) < \mathbf{0}$. $\square$

- If $h : \mathbb{R}^n \to \mathbb{R}^r$ is continuous and satisfies certain other conditions then:
  - the **boundary** of $\mathbb{S} = \{x \in \mathbb{R}^n | h(x) \leq \mathbf{0}\}$ is the set $\{x \in \mathbb{R}^n | h(x) \leq \mathbf{0}$ and, for at least one ℓ, $h_\ell(x) = 0\}$, and
  - its **interior** is the set $\{x \in \mathbb{R}^n | h(x) < \mathbf{0}\}$.
- That is, the set of strictly feasible points for the inequality constraints is the interior of $\mathbb{S}$.

Example

$$\forall x \in \mathbb{R}^2, h(x) = \begin{bmatrix} 3 - x_2 \\ x_1 + x_2 - 10 \end{bmatrix}.$$

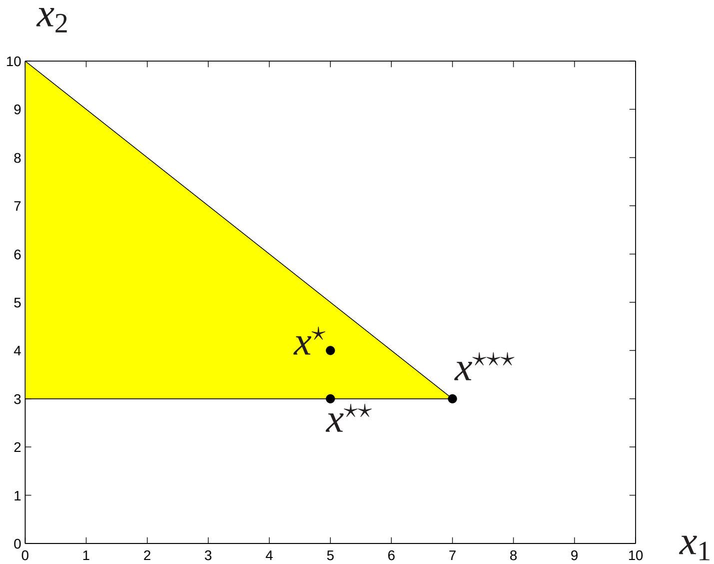


Fig. 2.13. Points $x^\star, x^{\star\star}$, and $x^{\star\star\star}$ that are feasible with respect to inequality constraints. The feasible set is the shaded triangular region for which $x_2 \geq 3$ and $x_1 + x_2 \leq 10$.

Example, continued

$$x^\star = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$$

- The constraints $h_1(x) \leq 0$ and $h_2(x) \leq 0$ are non-binding so that the active set is $\mathbb{A}(x^\star) = \emptyset$.
- This point is in the interior of the set $\{x \in \mathbb{R}^2 | h(x) \leq \mathbf{0}\}$.

$$x^{\star\star} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

- The constraint $h_2(x) \leq 0$ is non-binding while the constraint $h_1(x) \leq 0$ is binding so that the active set is $\mathbb{A}(x^{\star\star}) = \{1\}$.
- This point is on the boundary of the set $\{x \in \mathbb{R}^2 | h(x) \leq \mathbf{0}\}$.

$$x^{\star\star\star} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

- The constraints $h_1(x) \leq 0$ and $h_2(x) \leq 0$ are both binding so that the active set is $\mathbb{A}(x^{\star\star\star}) = \{1, 2\}$.
- This point is on the boundary of the set $\{x \in \mathbb{R}^2 | h(x) \leq \mathbf{0}\}$.

Example in higher dimension

- Consider Figure 2.14, which shows a **dodecahedron**, a twelve-sided solid, in $\mathbb{R}^3$.
- The dodecahedron is an example of a set that can be described in the form $\mathbb{S} = \{x \in \mathbb{R}^3 | h(x) \leq \mathbf{0}\}$ with $h : \mathbb{R}^3 \to \mathbb{R}^{12}$ affine.

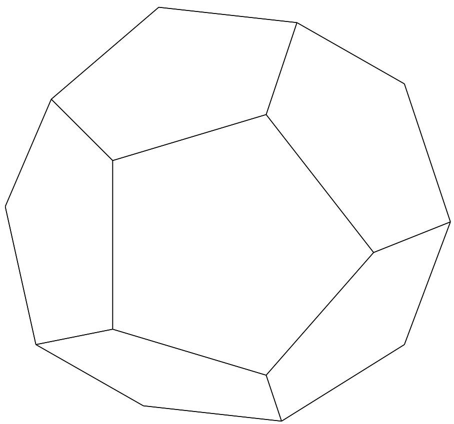


Fig. 2.14. Dodecahedron in $\mathbb{R}^3$.

Various cases for a point in $\mathbb{S}$

$x^\star$ is in the interior of the dodecahedron.

- We have $h(x^\star) < \mathbf{0}$ and $\mathbb{A}(x^\star) = \emptyset$.

$x^{\star\star}$ is on a face of the dodecahedron but not on an edge or vertex.

- That is, exactly one constraint ℓ is binding and $\mathbb{A}(x^{\star\star}) = \{\ell\}$.
- $x^{\star\star}$ is on the boundary.

$x^{\star\star\star}$ is on an edge but not a vertex of the dodecahedron.

- That is, exactly two constraints ℓ, ℓ' are binding and $\mathbb{A}(x^{\star\star\star}) = \{\ell, \ell'\}$.
- $x^{\star\star\star}$ is on the boundary.

$x^{\star\star\star\star}$ is a vertex of the dodecahedron.

- That is, exactly three constraints ℓ, ℓ', and ℓ'' are binding and $\mathbb{A}(x^{\star\star\star\star}) = \{\ell, \ell', \ell''\}$.
- $x^{\star\star\star\star}$ is on the boundary.

Discussion

- The importance of the notion of binding constraints is that it is typical for some but not all of the inequality constraints to be binding at the optimum.

Representation of inequality constraints

- Most optimization software can deal directly with:
  - **double-sided functional inequalities** such as $\underline{h} \leq h(x) \leq \overline{h}$ and
  - **double-sided inequalities on variables** such as $\underline{x} \leq x \leq \overline{x}$,
- For notational simplicity, we will usually restrict ourselves to inequalities of the form $h(x) \leq \mathbf{0}$, but recognize that problems may be easier to express in terms of the more comprehensive form $\underline{x} \leq x \leq \overline{x}, \underline{h} \leq h(x) \leq \overline{h}$.
- It is almost always worthwhile to take advantage of the more comprehensive form when the software has the capability.

Case studies

- Least-cost production with capacity constraints (in Section 15.1),
- Optimal routing in a data communications network (in Section 15.2),
- Least absolute value data fitting (in Section 15.3),
- Optimal margin pattern classification (in Section 15.4),
- Sizing of gate interconnects in integrated circuits (in Section 15.5), and
- Optimal power flow (in Section 15.6).

2.3.2.4 *Summary*

- For small example problems, inspection of a carefully drawn diagram can yield the minimum and minimizer.
- For larger problems where the dimension of x increases significantly past two, or the dimension of g or h increases, the geometry becomes more difficult to visualize and intuition becomes less reliable in predicting the solution.

2.3.3 Problems without minimum and the infimum

2.3.3.1 Analysis

- To discuss problems that do not have a minimum, we need a more general definition.

Definition 2.6 Let $\mathbb{S} \subseteq \mathbb{R}^n$, $f : \mathbb{S} \to \mathbb{R}$. Then, $\inf_{x\in\mathbb{S}} f(x)$, the **infimum** of the corresponding minimization problem, $\min_{x\in\mathbb{S}} f(x)$, is defined by:

$$\inf_{x\in\mathbb{S}} f(x) = \begin{cases} \text{the greatest lower bound for} & \\ \qquad\qquad \min_{x\in\mathbb{S}} f(x), & \text{if } \min_{x\in\mathbb{S}} f(x) \text{ is bounded below,} \\ -\infty, & \text{if } \min_{x\in\mathbb{S}} f(x) \text{ is unbounded below,} \\ \infty, & \text{if } \min_{x\in\mathbb{S}} f(x) \text{ is infeasible.} \end{cases}$$

By definition, the infimum is equal to the minimum of the corresponding minimization problem $\min_{x\in\mathbb{S}} f(x)$ if the minimum exists, but the infimum exists even if the problem has no minimum. To emphasize the role of $\mathbb{S}$, we also use the notation $\inf_{x\in\mathbb{R}^n}\{f(x)|x \in \mathbb{S}\}$ and analogous notations for the infimum. $\square$

2.3.3.2 Examples

Unconstrained problem with unbounded objective

$$\forall x \in \mathbb{R}, f(x) = x. \tag{2.20}$$

- There is no $f^\star \in \mathbb{R}$ such that $\forall x \in \mathbb{R}, f^\star \leq f(x)$.
- The problem $\min_{x \in \mathbb{R}} f(x)$ is unbounded below.
- The infimum is $\inf_{x \in \mathbb{R}} f(x) = -\infty$.

Unconstrained problem with objective that is bounded below

- $\underline{f} = 0$ is a lower bound for $\min_{x \in \mathbb{R}} f(x)$, where $\forall x \in \mathbb{R}, f(x) = \exp(x)$.
- The problem has no minimum but the infimum is $\inf_{x \in \mathbb{R}} f(x) = 0$.

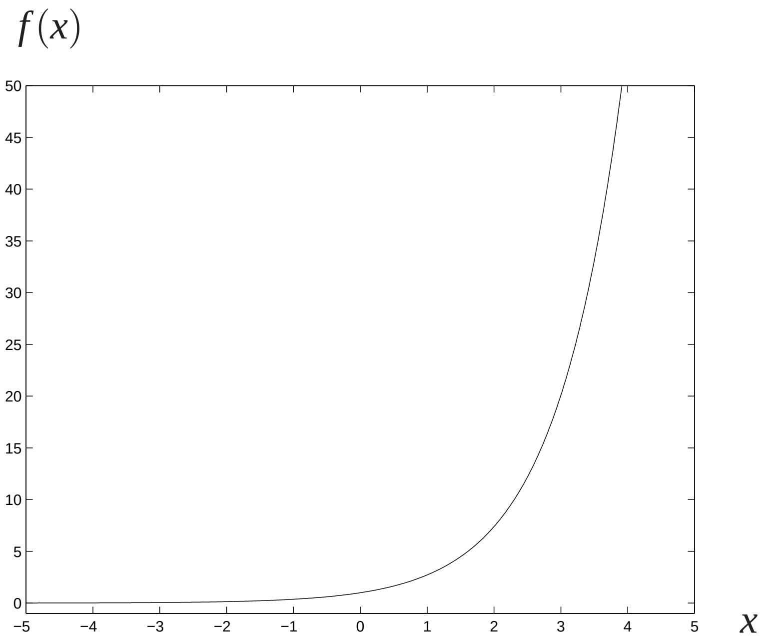


Fig. 2.15. The function exp is bounded below on the feasible set $\mathbb{R}$ but has no minimum.

Strict inequalities

- Again consider the objective $f : \mathbb{R} \to \mathbb{R}$ defined in (2.20):

$$\forall x \in \mathbb{R}, f(x) = x,$$

- but let the feasible set be:

$$\mathbb{S} = \{x \in \mathbb{R} | x > 0\}.$$

- Figure 2.16 shows the objective on the feasible set.
- Note that $\forall x \in \mathbb{S}, f(x) \geq 0$, so that the problem is bounded below by 0.
- However, there is no $x^\star \in \mathbb{S}$ such that $f(x^\star) = 0$.
- In this example, the problem is bounded, but does not have a minimum nor a minimizer.
- For this problem, the infimum is $\inf_{x \in \mathbb{R}} \{f(x) | x > 0\} = 0$.

Strict inequalities, continued

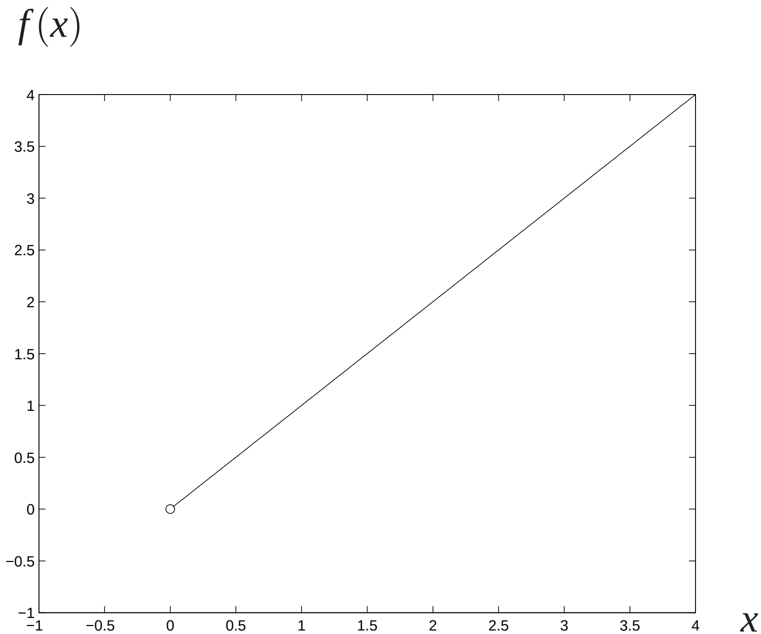


Fig. 2.16. Function that is bounded below on feasible set but where the problem has no minimum because the feasible set is defined by a strict inequality. The function is illustrated only on the feasible set. The circle $\circ$ at $x = 0$, $f(x) = 0$ indicates that this point is not included in the graph but that points to the right of $x = 0$ and arbitrarily close to $x = 0$ are included in the graph.

Inconsistent constraints

- Consider any objective $f : \mathbb{R} \to \mathbb{R}$ and let:

$$\mathbb{S} = \{x \in \mathbb{R} | g(x) = \mathbf{0}\},$$

- where $g : \mathbb{R} \to \mathbb{R}^2$ is defined by:

$$\forall x \in \mathbb{R}, g(x) = \begin{bmatrix} x+1 \\ x-1 \end{bmatrix}.$$

- Then there are no feasible solutions, since the equality constraints are **inconsistent** and so $\mathbb{S} = \emptyset$.
- In this example, there are no feasible values of x and therefore no minimum.
- The infimum is $\inf_{x \in \mathbb{R}} \{f(x) | g(x) = \mathbf{0}\} = \infty$.

Discontinuous objective

- Finally, let:

$$\mathbb{S} = \{x \in \mathbb{R} | x \geq 0\},$$

- and define $f : \mathbb{S} \to \mathbb{R}$ by:

$$\forall x \in \mathbb{S}, f(x) = \begin{cases} 1, & \text{if } x = 0, \\ x, & \text{if } x \neq 0. \end{cases} \tag{2.21}$$

Discontinuous objective, continued

- The problem $\min_{x \in \mathbb{S}} f(x)$ is bounded below by zero, but there is again no minimum nor minimizer.
- The infimum is $\inf_{x \in \mathbb{R}} \{f(x) | x \geq 0\} = 0$.

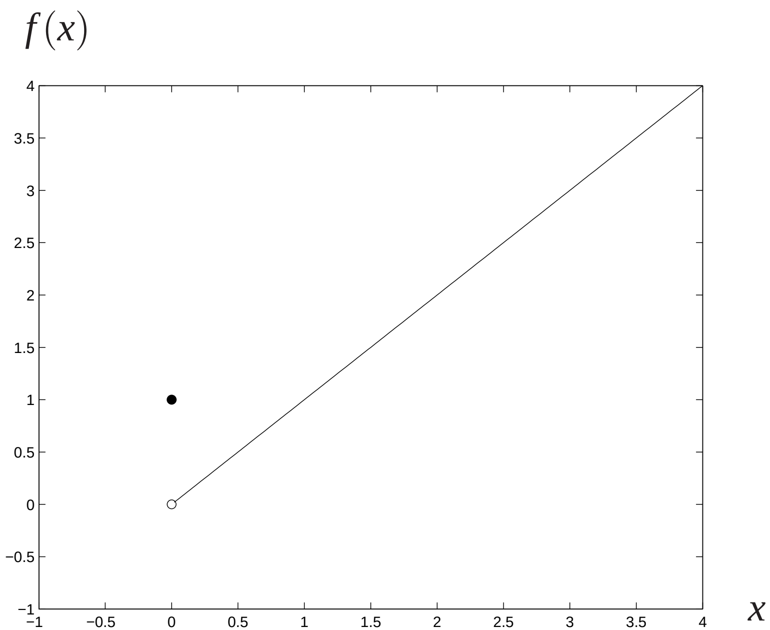


Fig. 2.17. Function (2.21) that is bounded below on feasible set but where the problem has no minimum because the function is discontinuous. The function is illustrated only on the feasible set. The bullet • at $x = 0$, $f(x) = 1$ indicates that this is the value of the function at $x = 0$.

2.3.3.3 Summary

- In all five cases, $\operatorname{argmin}_{x\in\mathbb{R}} f(x)$ is the empty set $\emptyset$.
- Careful formulation of a problem can avoid these issues.

2.3.4 Conditions for problems to possess a minimum and minimizer

Theorem 2.1 *Let $\mathbb{S} \subseteq \mathbb{R}^n$ be non-empty, closed, and bounded and let $f : \mathbb{S} \to \mathbb{R}$ be continuous. Then the problem $\min_{x\in\mathbb{S}} f(x)$ possesses a minimum and minimizer.* $\square$

2.3.5 *Maximization problems and the supremum*

$$\max_{x \in \mathbb{S}} f(x) = -\min_{x \in \mathbb{S}}(-f(x)). \tag{2.22}$$

Definition 2.7 Let $\mathbb{S} \subseteq \mathbb{R}^n$, $f : \mathbb{S} \to \mathbb{R}$. Then, $\sup_{x \in \mathbb{S}} f(x)$, the **supremum** of the corresponding maximization problem $\max_{x \in \mathbb{S}} f(x)$ is defined by:

$$\sup_{x \in \mathbb{S}} f(x) = \begin{cases} \text{the } \textbf{least upper bound} \text{ for} & \\ \qquad \max_{x \in \mathbb{S}} f(x), & \text{if } \max_{x \in \mathbb{S}} f(x) \text{ is bounded above,} \\ \infty, & \text{if } \max_{x \in \mathbb{S}} f(x) \text{ is unbounded above,} \\ -\infty, & \text{if } \max_{x \in \mathbb{S}} f(x) \text{ is infeasible.} \end{cases}$$

The supremum is equal to the maximum of the corresponding maximization problem $\max_{x \in \mathbb{S}} f(x)$ if the maximum exists. $\square$

2.3.6 Extended real functions

Definition 2.8 Let $\mathbb{S} \subseteq \mathbb{R}^n$, $f : \mathbb{S} \to \mathbb{R} \cup \{-\infty, \infty\}$, and $f^\star \in \mathbb{R}$. Then by:

$$f^\star = \min_{x \in \mathbb{S}} f(x),$$

we mean that:

$$\exists x^\star \in \mathbb{S} \text{ such that } f^\star = f(x^\star) \in \mathbb{R} \text{ and } (x \in \mathbb{S}) \Rightarrow (f(x^\star) \leq f(x)).$$

$\square$

2.4 Algorithms

- Two basic types of algorithms:
  - Direct, to be described in Section 2.4.1, and
  - Iterative, to be described in Section 2.4.2.

2.4.1 Direct

- A finite list of **operations** that calculates the solution of the problem.

2.4.1.1 Discussion

- Under the (usually unrealistic) assumptions that:
  - all numbers in the problem specification are represented to **infinite precision**,
  - all arithmetic operations are carried out to infinite precision, and
  - the answers to each arithmetic operation are represented to infinite precision,
- then the answer obtained from a direct algorithm would be exact.

2.4.1.2 Applicability

- Some problems *cannot* be solved by direct algorithms.
- Consider $g : \mathbb{R} \to \mathbb{R}$ such that g is a **polynomial**.
- For non-linear equations involving arbitrary fifth or higher degree polynomials, there is *provably* no direct algorithm available to find the solution.

2.4.2 Iterative

2.4.2.1 Recursion to define iterates

- $x^{(\nu+1)} = x^{(\nu)} + \alpha^{(\nu)} \Delta x^{(\nu)}, \nu = 0, 1, 2, \ldots,$ is the iteration counter,
- $x^{(0)}$ is the **initial guess** of the solution,
- $x^{(\nu)}$ is the value of the iterate at the ν-th iteration,
- $\alpha^{(\nu)} \in \mathbb{R}_+$ is the **step-size**, with usually $0 < \alpha^{(\nu)} \leq 1$,
- $\Delta x^{(\nu)} \in \mathbb{R}^n$ is the **step direction**, and
- the product $\alpha^{(\nu)} \Delta x^{(\nu)}$ is the **update** to add to the current iterate $x^{(\nu)}$.

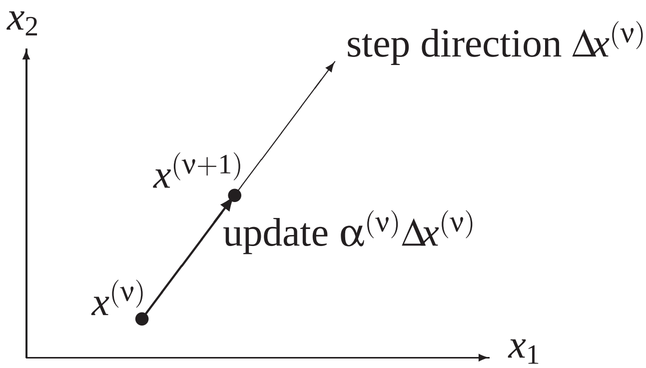


Fig. 2.18. Update of iterate in $\mathbb{R}^2$. The bullets • indicate the locations of the points $x^{(\nu)}$ and $x^{(\nu+1)}$, while the arrows $\nearrow$ indicate the magnitudes and directions of the vectors $\Delta x^{(\nu)}$ and $\alpha^{(\nu)} \Delta x^{(\nu)}$.

2.4.2.2 Sequence of iterates and closeness to a solution

Definition 2.9 Let $\|\bullet\|$ be a norm on $\mathbb{R}^n$. Let $\{x^{(\nu)}\}_{\nu=0}^{\infty}$ be a sequence of vectors in $\mathbb{R}^n$. Then, the sequence $\{x^{(\nu)}\}_{\nu=0}^{\infty}$ **converges** to a **limit** $x^\star$ if:

$$\forall \varepsilon > 0, \exists N^\varepsilon \in \mathbb{Z}_+ \text{ such that } (\nu \in \mathbb{Z}_+ \text{ and } \nu \geq N^\varepsilon) \Rightarrow \left(\left\| x^{(\nu)} - x^\star \right\| \leq \varepsilon \right).$$

The set $\mathbb{Z}_+$ is the set of non-negative integers.
If the sequence $\{x^{(\nu)}\}_{\nu=0}^{\infty}$ converges to $x^\star$ then we write $\lim_{\nu\to\infty} x^{(\nu)} = x^\star$ or $\lim\limits_{\nu\to\infty} x^{(\nu)} = x^\star$ and call $x^\star$ the **limit** of the sequence $\{x^{(\nu)}\}_{\nu=0}^{\infty}$. $\square$

Analysis

Definition 2.10 Let $\|\bullet\|$ be a norm. A sequence $\{x^{(\nu)}\}_{\nu=0}^{\infty}$ that converges to $x^\star \in \mathbb{R}^n$ is said to converge at **rate** $R \in \mathbb{R}_{++}$ (where $\mathbb{R}_{++}$ is the set of strictly positive real numbers) and with **rate constant** $C \in \mathbb{R}_{++}$ if:

$$\lim_{\nu\to\infty} \frac{\left\|x^{(\nu+1)} - x^\star\right\|}{\left\|x^{(\nu)} - x^\star\right\|^R} = C. \tag{2.23}$$

If (2.23) is satisfied for $R = 1$ and some value of C in the range $0 < C < 1$ then the rate is called **linear**. If (2.23) is satisfied for $R = 2$ and some C in the range $0 < C < \infty$ then the rate is called **quadratic**. If (2.23) is satisfied for some R in the range $1 < R < 2$ and some C in the range $0 < C < \infty$ then the rate is called **super-linear**. □

Discussion

- Qualitatively, the larger the value of R, the faster the iterates converge, at least asymptotically.

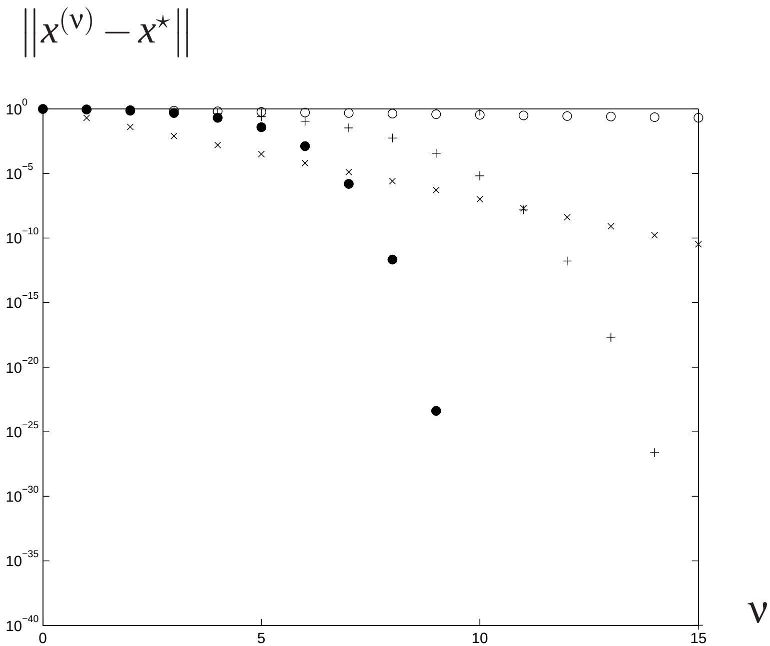


Fig. 2.19. Rates of convergence for several sequences, with: $R = 1$ and $C = 0.9$ shown as $\circ$; $R = 1$ and $C = 0.2$ shown as $\times$; $R = 2$ and $C = 0.9$ shown as $\bullet$; and $R = 1.5$ and $C = 0.9$ shown as $+$.

2.5 Solutions of simultaneous equations

2.5.1 Number of solutions

- Consider a linear equation in one variable, $Ax = b$, where $A, b \in \mathbb{R}$.
- The possible cases are:

$$\begin{aligned} 0x &= 0, && \text{infinitely many solutions,} \\ 0x &= b, && b \neq 0,\ \text{no solutions,} \\ Ax &= b, && A \neq 0,\ \text{one solution.} \end{aligned}$$

2.5.2 Uniqueness of solution for linear equations

- Necessary and sufficient conditions for there to be a unique solution to a square system of equations is that the coefficient matrix A be **non-singular**.

Number of solutions, continued

- Consider a quadratic equation in one variable, $Q(x)^2 + Ax = b.$ where $A, b, Q \in \mathbb{R}$.
- The possible cases are:

$$
\begin{aligned}
0(x)^2 + 0x = 0, &\quad \text{infinitely many solutions,} \\
0(x)^2 + 0x = b, &\quad b \neq 0, \text{ no solutions,} \\
0(x)^2 + Ax = b, &\quad A \neq 0, \text{ one solution,} \\
Q(x)^2 + Ax = b, &\quad Q \neq 0, A^2 + 4Qb < 0, \text{ no (real) solutions,} \\
Q(x)^2 + Ax = b, &\quad Q \neq 0, A^2 + 4Qb = 0, \text{ one solution,} \\
Q(x)^2 + Ax = b, &\quad Q \neq 0, A^2 + 4Qb > 0, \text{ two solutions.}
\end{aligned}
$$

2.5.3 Uniqueness of solution for non-linear equations

- To study uniqueness, we will consider simultaneous equations where the number of equations equals the number of variables.

2.5.3.1 Monotone functions

Definition 2.11 Let $\mathbb{S} \subseteq \mathbb{R}^n$ and let $g : \mathbb{S} \rightarrow \mathbb{R}^n$. We say that g is **monotone** on $\mathbb{S}$ if:

$$\forall x, x' \in \mathbb{S}, \left(g(x') - g(x)\right)^\dagger (x' - x) \geq 0. \tag{2.24}$$

We say that g is **strictly monotone** on $\mathbb{S}$ if:

$$\forall x, x' \in \mathbb{S}, (x \neq x') \Rightarrow \left(g(x') - g(x)\right)^\dagger (x' - x) > 0.$$

If g is monotone on $\mathbb{R}^n$ then we say that g is monotone. If g is strictly monotone on $\mathbb{R}^n$ then we say that g is strictly monotone. $\square$

Monotone functions, continued

- Geometrically, g is monotone on $\mathbb{S}$ if, for all pairs of vectors x and x' in $\mathbb{S}$, the vectors $(x' - x)$ and $(g(x') - g(x))$ point in directions that are within less than or equal to 90° of each other.

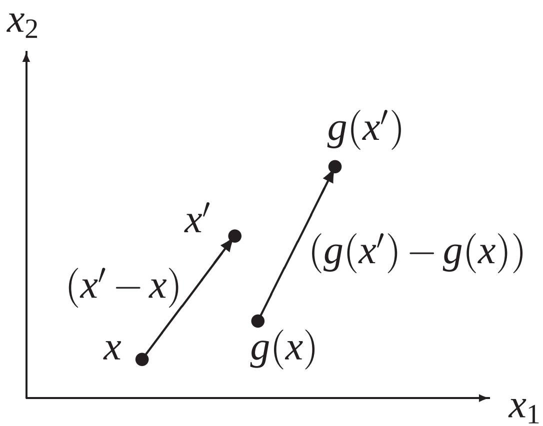


Fig. 2.20. Illustration of definition of monotone. For all x and x' in $\mathbb{S}$, the vectors $(x' - x)$ and $(g(x') - g(x))$ point in directions that are within less than or equal to 90° of each other.

Example

- Even if a function $\hat{g} : \mathbb{R}^n \to \mathbb{R}^n$ is not strictly monotone, by permuting the entries of $\hat{g}$ it may be possible to create a strictly monotone function.
- Consider the function:

$$\forall x \in \mathbb{R}^2, \hat{g}(x) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}.$$

- This function is not strictly monotone since:

$$\begin{aligned} \left(\hat{g}(x') - \hat{g}(x)\right)^{\dagger}(x' - x) &= 2(x'_2 - x_2)(x'_1 - x_1), \\ &< 0, \text{ if } x'_2 > x_2 \text{ and } x'_1 < x_1. \end{aligned}$$

- However, the function $g : \mathbb{R}^2 \to \mathbb{R}^2$ obtained by swapping the entries of $\hat{g}$ is strictly monotone, since:

$$\begin{aligned} \left(g(x') - g(x)\right)^{\dagger}(x' - x) &= \left\|x' - x\right\|_2^2, \\ &> 0, \text{ for } x' \neq x. \end{aligned}$$

Analysis

Theorem 2.2 *Let $\mathbb{S} \subseteq \mathbb{R}^n$ and $g : \mathbb{S} \to \mathbb{R}^n$ be strictly monotone on $\mathbb{S}$. Then there is at most one solution of the simultaneous equations $g(x) = \mathbf{0}$ that is an element of $\mathbb{S}$.*

Proof Suppose that there are two solutions $x^\star, x^{\star\star} \in \mathbb{S}$ with $x^\star \neq x^{\star\star}$. That is, $g(x^\star) = g(x^{\star\star}) = \mathbf{0}$. Consequently, $(g(x^\star) - g(x^{\star\star}))^\dagger (x^\star - x^{\star\star}) = 0$. But by definition of strictly monotone applied to $x^\star$ and $x^{\star\star}$, $(g(x^\star) - g(x^{\star\star}))^\dagger (x^\star - x^{\star\star}) > 0$. This is a contradiction. $\square$

Discussion

- It is possible for a function g to be not strictly monotone and yet there may be a unique solution or no solution to the equations $g(x) = \mathbf{0}$.

Example

- Consider $g : \mathbb{R} \to \mathbb{R}$ defined by $\forall x \in \mathbb{R}, g(x) = (x)^3 - x - 6$.
- This function is not strictly monotone, yet there is only one solution to $g(x) = 0$, namely $x^\star = 2$.

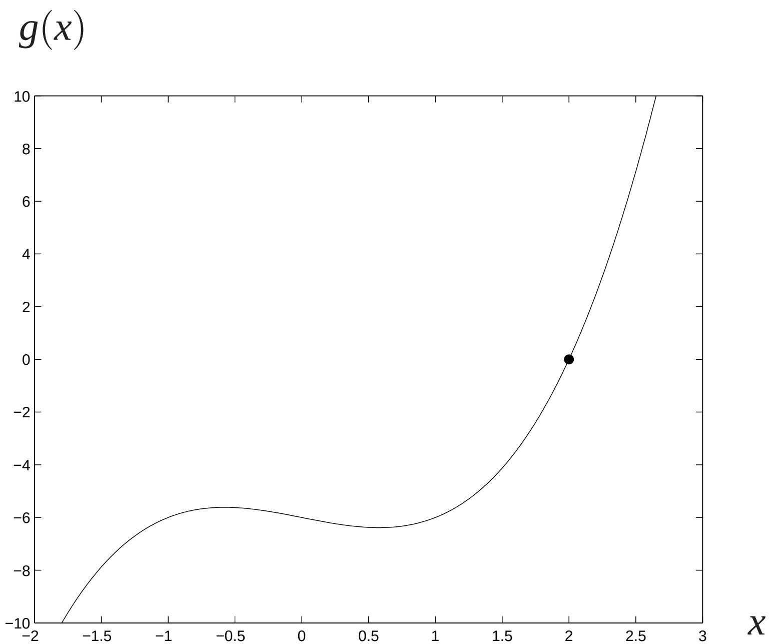


Fig. 2.21. Function g that is not strictly monotone but for which there is only one solution, $x^\star = 2$, to $g(x) = 0$. The solution is illustrated with the $\bullet$.

2.5.3.2 *Characterizing monotone and strictly monotone functions*

Jacobian

- The entries of the **Jacobian** $J : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ are defined by:

$$\forall k = 1, \ldots, n, \forall \ell = 1, \ldots, m, J_{\ell k} = \frac{\partial g_\ell}{\partial x_k}.$$

Positive definite and positive semi-definite

- A matrix $Q \in \mathbb{R}^{n \times n}$ is **positive semi-definite** if:

$$\forall x \in \mathbb{R}^n, x^\dagger Q x \geq 0.$$

- The matrix is **positive definite** if:

$$\forall x \in \mathbb{R}^n, (x \neq \mathbf{0}) \Rightarrow (x^\dagger Q x > 0).$$

Convex sets

Definition 2.12 Let $\mathbb{S} \subseteq \mathbb{R}^n$. We say that $\mathbb{S}$ is a **convex set** or that $\mathbb{S}$ is **convex** if $\forall x, x' \in \mathbb{S}, \forall t \in [0,1], (1-t)x + tx' \in \mathbb{S}$. $\square$

- The sets $\mathbb{R}^n$, $\mathbb{R}^n_+$, and $\mathbb{R}^n_{++}$ are all convex.

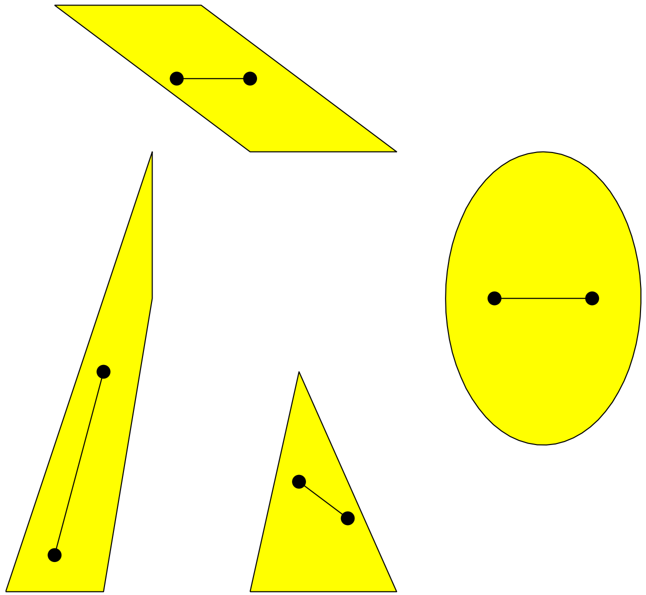


Fig. 2.22. Convex sets with pairs of points joined by line segments.

Examples of non-convex sets

- Non-convex sets can have “indentations.”

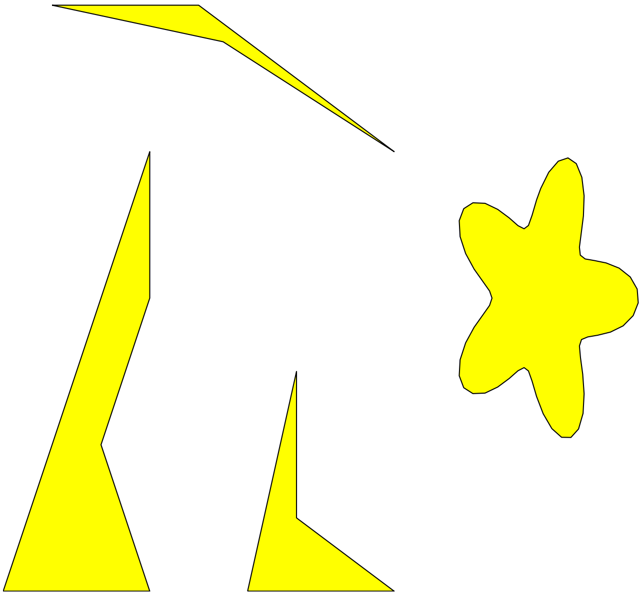


Fig. 2.23. Non-convex sets.

Conditions for strictly monotone

Theorem 2.3 *Let $\mathbb{S} \subseteq \mathbb{R}^n$ be a convex set and $g : \mathbb{S} \to \mathbb{R}^n$. Suppose that g is partially differentiable with continuous partial derivatives on $\mathbb{S}$. Moreover, suppose that the Jacobian J is positive semi-definite throughout $\mathbb{S}$. Then g is monotone on $\mathbb{S}$. If J is positive definite throughout $\mathbb{S}$ then g is strictly monotone on $\mathbb{S}$.*

Proof Suppose that J is positive semi-definite throughout $\mathbb{S}$. Let $x, x' \in \mathbb{S}$. For $0 \leq t \leq 1$ we have that $(x + t[x' - x]) \in \mathbb{S}$ since $\mathbb{S}$ is a convex set. As t varies from 0 to 1, $(x + t[x' - x])$ traces out the line segment joining x and x'. Define $\phi : [0,1] \to \mathbb{R}$ by:

$$\begin{aligned} \forall t \in [0,1], \phi(t) &= (x' - x)^\dagger g(x + t[x' - x]), \\ &= g(x + t[x' - x])^\dagger (x' - x). \end{aligned}$$

Proof, continued We have:

$$\begin{aligned}
\phi(1)-\phi(0) &= (x'-x)^{\dagger}(g(x')-g(x)), \\
&= (g(x')-g(x))^{\dagger}(x'-x),
\end{aligned}$$

and so we must prove that $\phi(1)-\phi(0)\geq 0$. Notice that:

$$\begin{aligned}
\frac{d\phi}{dt}(t) &= (x'-x)^{\dagger}J(x+t[x'-x])(x'-x), \text{ by the chain rule,} \\
&\geq 0, \text{ for } 0\leq t\leq 1, \qquad (2.25)
\end{aligned}$$

since $J(x+t[x'-x])$ is positive semi-definite. We have:

$$\begin{aligned}
\phi(1) &= \phi(0)+\int_{t=0}^{1}\frac{d\phi}{dt}(t)\,dt, \\
&\qquad \text{by the fundamental theorem of calculus applied to } \phi, \\
&\geq \phi(0), \text{ since the integrand is non-negative everywhere by (2.25).}
\end{aligned}$$

This is the result we were trying to prove. A similar analysis applies for J positive definite, noting that the integrand is then strictly positive and continuous. □

2.6 Solutions of optimization problems

2.6.1 Local and global minima

2.6.1.1 Definitions

- Recall Problem (2.8) and its minimum $f^\star$:

$$f^\star = \min_{x \in \mathbb{S}} f(x).$$

- Sometimes, we call $f^\star$ in Problem (2.8) the **global** minimum of the problem to emphasize that there is no $x \in \mathbb{S}$ that has a smaller value of $f(x)$.

Definition 2.13 Let $\|\bullet\|$ be a norm on $\mathbb{R}^n$, $\mathbb{S} \subseteq \mathbb{R}^n$, $x^\star \in \mathbb{S}$, and $f : \mathbb{S} \to \mathbb{R}$. We say that $x^\star$ is a **local minimizer** of the problem $\min_{x \in \mathbb{S}} f(x)$ if:

$$\exists \varepsilon > 0 \text{ such that } \forall x \in \mathbb{S}, (\|x - x^\star\| < \varepsilon) \Rightarrow (f(x^\star) \leq f(x)). \tag{2.26}$$

The value $f^\star = f(x^\star)$ is called a **local minimum** of the problem. □

Local minimizer and minimum

- A local minimum may or may not be a global minimum but if a problem possesses a minimum then there is exactly one global minimum, by definition.
- The global minimum is also a local minimum.
- Formally, $\hat{x}$ is not a local minimizer if:

$$\forall \varepsilon > 0, \exists x^{\varepsilon} \in \mathbb{S} \text{ such that } (\|\hat{x} - x^{\varepsilon}\| < \varepsilon) \text{ and } (f(\hat{x}) > f(x^{\varepsilon})). \tag{2.27}$$

2.6.1.2 Examples

Multiple local minimizers over a convex set

- $f : \mathbb{R} \to \mathbb{R}$ has two local minimizers at $x^\star = 3, x^{\star\star} = -3$ over $\mathbb{S}$.

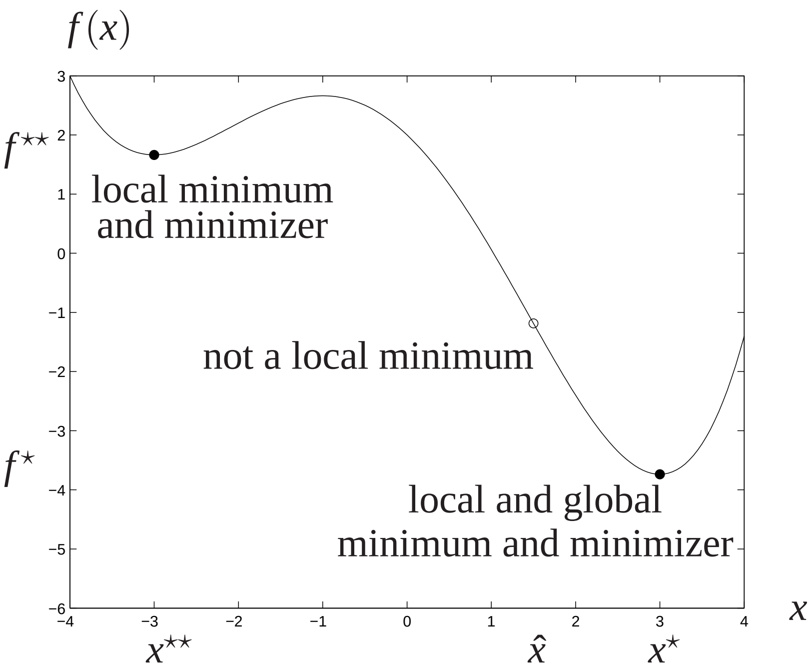


Fig. 2.24. Local minima, $f^\star$ and $f^{\star\star}$, with corresponding local minimizers $x^\star$ and $x^{\star\star}$, over a set $\mathbb{S}$. The point $x^\star$ is the global minimizer and $f^\star$ the global minimum over $\mathbb{S}$.

Illustration of definition of not a local minimizer

- For $\varepsilon = 1$ there is a point, namely $\hat{x} + \frac{\varepsilon}{2} = 2$ that is within a distance ε of $\hat{x} = 1.5$ and which has lower value of objective than the point $\hat{x} = 1.5$.

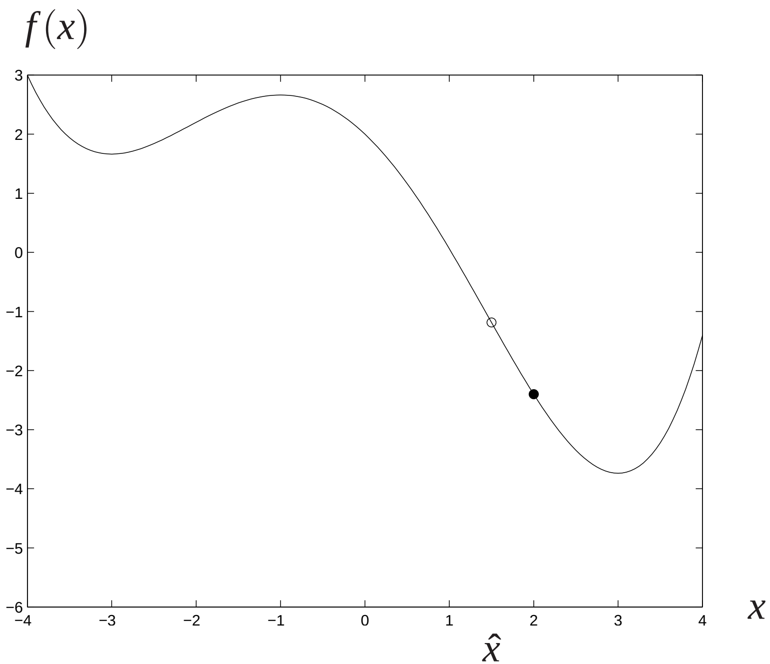


Fig. 2.25. A point $\hat{x} = 1.5$, illustrated with a $\circ$, that is not a local minimizer and another point, $\hat{x} + \frac{\varepsilon}{2} = 2$, illustrated with a $\bullet$, that is within a distance $\varepsilon = 1$ of $\hat{x}$ and has a lower objective value.

Illustration of definition of not a local minimizer, continued

- For $\varepsilon = 0.5$ there is a point, namely $\hat{x} + \frac{\varepsilon}{2} = 1.75$ that is within a distance ε of $\hat{x} = 1.5$ and which has lower value of objective than the point $\hat{x} = 1.5$.

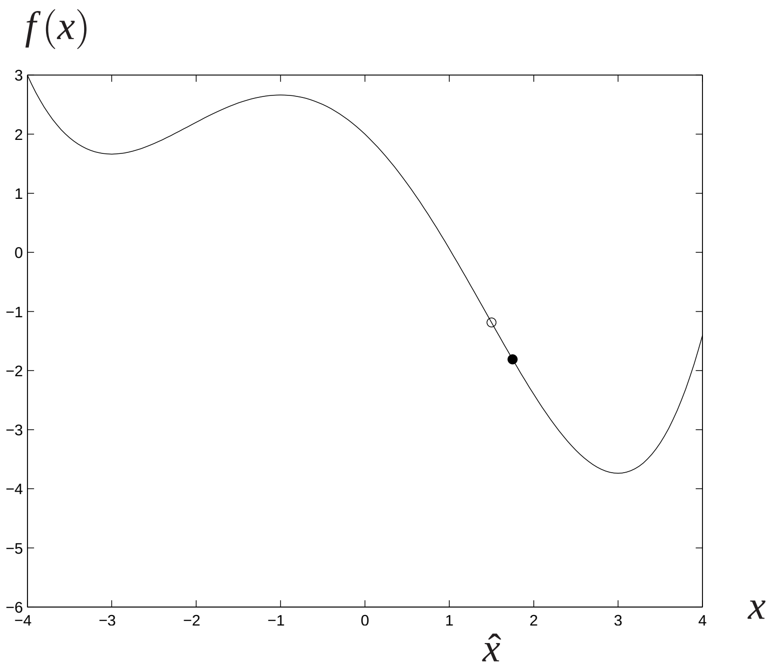


Fig. 2.26. A point $\hat{x} = 1.5$, illustrated with a $\circ$, that is not a local minimizer and another point, $\hat{x} + \frac{\varepsilon}{2} = 1.75$, illustrated with a $\bullet$, that is within a distance $\varepsilon = 0.5$ of $\hat{x}$ and has a lower objective value.

Multiple local minimizers over a non-convex set

- Over the non-convex set $\mathbb{P} = \{x \in \mathbb{R} | -4 \leq x \leq 1 \text{ or } 2 \leq x \leq 4\}$ there are three local minimizers, $x^\star = 3, x^{\star\star} = -3$, and $x^{\star\star\star} = 1$.

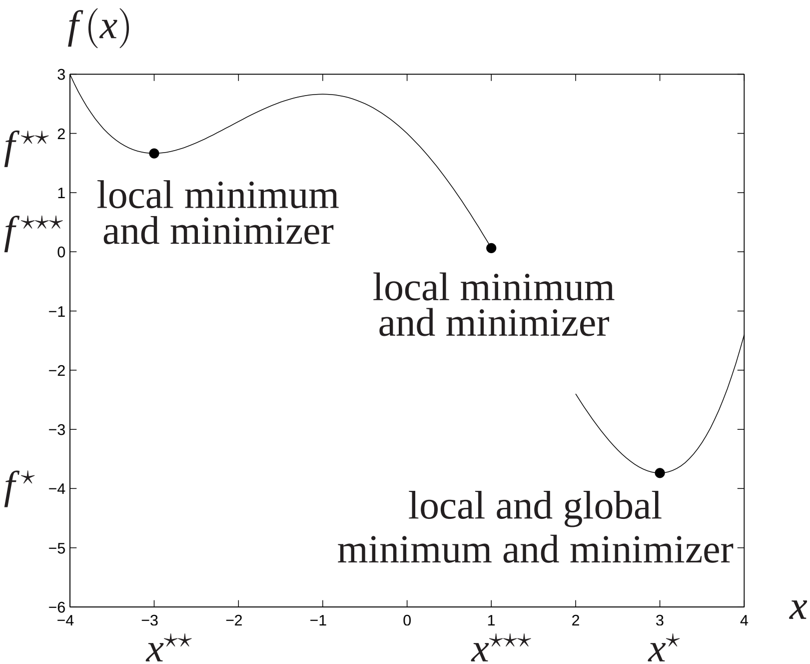


Fig. 2.27. Local and global minima and minimizers of a problem over a set $\mathbb{P} = \{x \in \mathbb{R} | -4 \leq x \leq 1 \text{ or } 2 \leq x \leq 4\}$.

Multiple local minimizers over a non-convex set in higher dimension

- The local minimizers are $x^\star \approx \begin{bmatrix} 2.4 \\ -0.1 \end{bmatrix}$ and $x^{\star\star} \approx \begin{bmatrix} 0.8 \\ -0.7 \end{bmatrix}$.

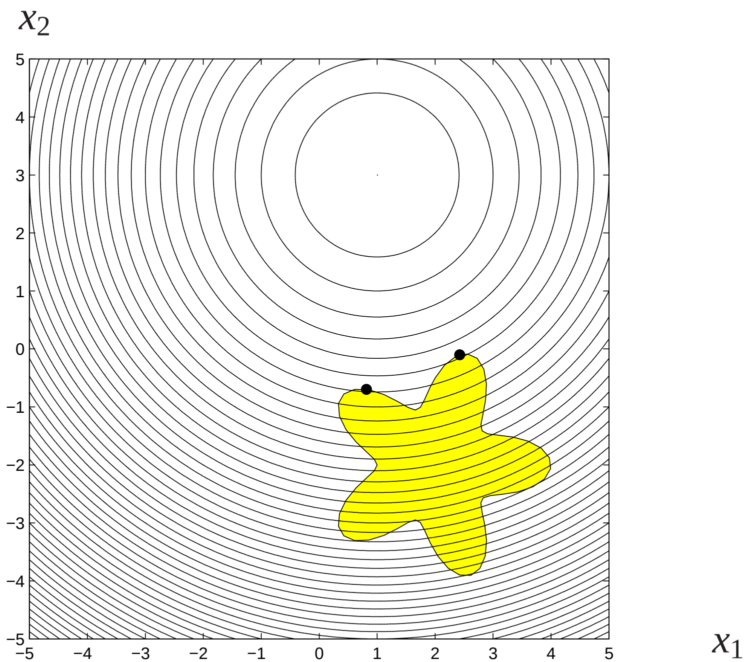


Fig. 2.28. Contour sets of the function defined in (2.10) with feasible set shaded. The two local minimizers are indicated by bullets. The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

2.6.1.3 Discussion

- Iterative algorithms involve generating a sequence of successively "better" points that provide successively better values of the objective or closer satisfaction of the constraints or both.
- With an iterative improvement algorithm, we can usually only guarantee, at best, that we are moving towards a local minimum and minimizer.

2.6.2 Strict and non-strict minimizers

2.6.2.1 Definitions

- There can be more than one minimizer even if the minimum is global.

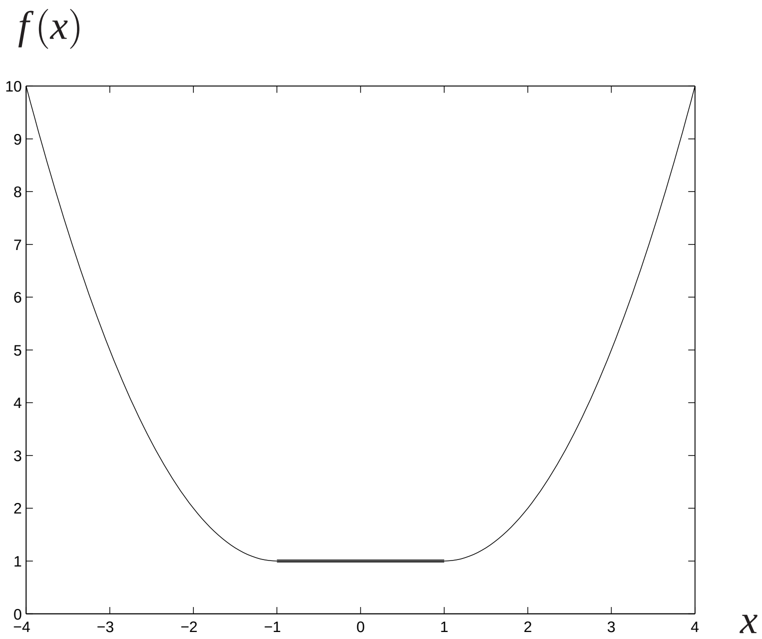


Fig. 2.29. A function with multiple global minimizers. The set of minimizers is indicated by a thick line.

Definition 2.14 We say that $x^\star \in \mathbb{S}$ is a **strict global** minimizer of the problem $\min_{x\in\mathbb{S}} f(x)$ if:

$$\forall x \in \mathbb{S}, (x \neq x^\star) \Rightarrow (f(x^\star) < f(x)).$$

The value $f^\star = f(x^\star)$ is called a **strict global** minimum of the problem. □

Definition 2.15 We say that $x^\star \in \mathbb{S}$ is a **strict local** minimizer of the problem $\min_{x\in\mathbb{S}} f(x)$ if:

$$\exists \varepsilon > 0 \text{ such that } \forall x \in \mathbb{S}, (0 < \|x - x^\star\| < \varepsilon) \Rightarrow (f(x^\star) < f(x)).$$

The value $f^\star = f(x^\star)$ is called a **strict local** minimum of the problem. □

2.6.2.2 Examples

- The two local minimizers, $x^\star = 3$ and $x^{\star\star} = -3$, in Figure 2.24 are strict local minimizers.
- All three local minimizers, $x^\star = 3, x^{\star\star} = -3, x^{\star\star\star} = 1$, in Figure 2.27 are strict local minimizers.

2.6.3 Convex functions

2.6.3.1 Definitions

Definition 2.16 Let $\mathbb{S} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathbb{S} \to \mathbb{R}$. Then, f is a **convex function** on $\mathbb{S}$ if:

$$\forall x, x' \in \mathbb{S}, \forall t \in [0,1], f([1-t]x + tx') \leq [1-t]f(x) + tf(x'). \tag{2.28}$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex on $\mathbb{R}^n$ then we say that f is convex. A function $h : \mathbb{S} \to \mathbb{R}^r$ is convex on $\mathbb{S}$ if each of its components h_ℓ is convex on $\mathbb{S}$. If $h : \mathbb{R}^n \to \mathbb{R}^r$ is convex on $\mathbb{R}^n$ then we say that h is convex. The set $\mathbb{S}$ is called the **test set**.

Furthermore, f is a **strictly convex function** on $\mathbb{S}$ if:

$$\forall x, x' \in \mathbb{S}, (x \neq x') \Rightarrow \left(\forall t \in (0,1), f([1-t]x + tx') < [1-t]f(x) + tf(x')\right).$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex on $\mathbb{R}^n$ then we say that f is strictly convex. A function $h : \mathbb{S} \to \mathbb{R}^r$ is strictly convex on $\mathbb{S}$ if each of its components h_ℓ is strictly convex on $\mathbb{S}$. If $h : \mathbb{R}^n \to \mathbb{R}^r$ is strictly convex on $\mathbb{R}^n$ then we say that h is strictly convex. $\square$

Discussion

- The condition in (2.28) means that linear interpolation of convex f between points on the curve is never below the function values.

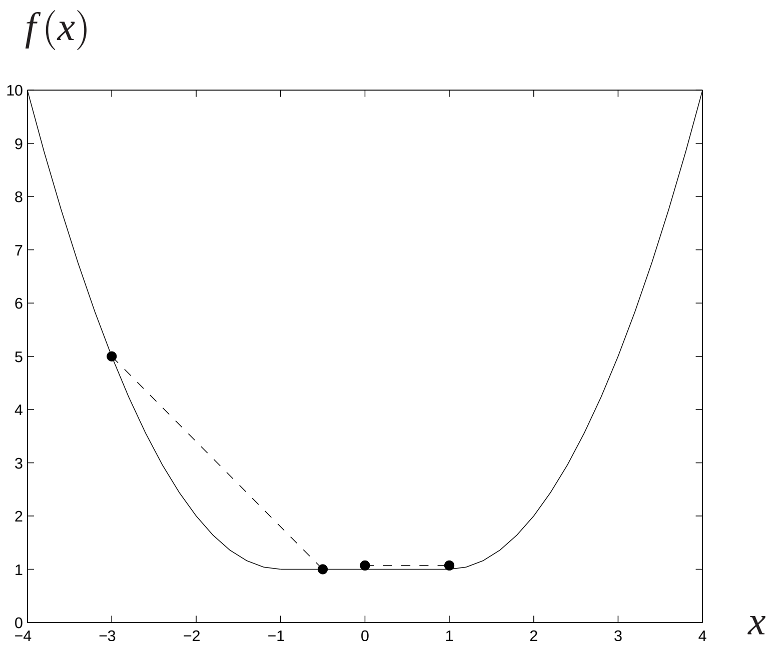


Fig. 2.30. Linear interpolation of a convex function between points never under-estimates the function. (For clarity, the line interpolating f between $x = 0$ and $x = 1$ is drawn slightly above the solid curve: it should be coincident with the solid curve.)

Definition 2.17 Let $\mathbb{S} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathbb{S} \to \mathbb{R}$. We say that f is a **concave function** on $\mathbb{S}$ if $(-f)$ is a convex function on $\mathbb{S}$. $\square$

2.6.3.2 Examples

- A linear or affine function is convex and concave on any convex set.
- The function $f : \mathbb{R} \to \mathbb{R}$ shown in Figure 2.24 is not convex on the convex set $\mathbb{S} = \{x \in \mathbb{R} | -4 \leq x \leq 4\}$.
- Qualitatively, convex functions are "bowl-shaped" and have level sets that are convex sets as specified in:

Definition 2.18 Let $\mathbb{S} \subseteq \mathbb{R}^n$ and $f : \mathbb{S} \to \mathbb{R}$. Then the function f has **convex level sets** on $\mathbb{S}$ if for all $\tilde{f} \in \mathbb{R}$ we have that $\mathbb{L}_f(\tilde{f})$ is convex. If $f : \mathbb{R}^n \to \mathbb{R}$ has convex level sets on $\mathbb{R}^n$ then we say that f has convex level sets. $\square$

- Note that a function with convex level sets need not itself be a convex function.

Convexity of level sets of convex function

$$\forall x \in \mathbb{R}^2, f(x) = (x_1 - 1)^2 + (x_2 - 3)^2 - 1.8(x_1 - 1)(x_2 - 3). \qquad (2.29)$$

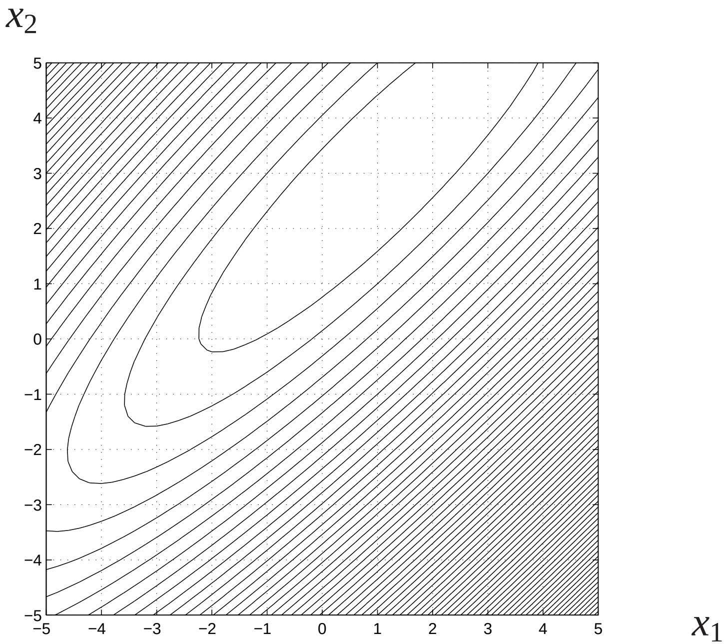


Fig. 2.31. Contour sets $\mathbb{C}_f(\tilde{f})$ of the function defined in (2.29). The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

2.6.3.3 Relationship to optimization problems

Theorem 2.4 *Let $\mathbb{S} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathbb{S} \to \mathbb{R}$. Then:*

(i) *If f is convex on $\mathbb{S}$ then it has at most one local minimum over $\mathbb{S}$.*

(ii) *If f is convex on $\mathbb{S}$ and has a local minimum over $\mathbb{S}$ then the local minimum is the global minimum.*

(iii) *If f is strictly convex on $\mathbb{S}$ then it has at most one minimizer over $\mathbb{S}$.*

Proof We prove all three items by contradiction.

(i) For the sake of a contradiction, suppose that f is convex, yet that it has two local minima over $\mathbb{S}$; that is, there are two distinct values $f^\star \in \mathbb{R}$ and $f^{\star\star} \in \mathbb{R}$, say, with $f^\star \neq f^{\star\star}$ that each satisfy Definition 2.13.
For concreteness, suppose that $f^\star > f^{\star\star}$ and let $x^\star \in \mathbb{S}$ and $x^{\star\star} \in \mathbb{S}$ be any two local minimizers associated with $f^\star$ and $f^{\star\star}$, respectively. The situation is illustrated in Figure 2.32.

Proof of (i), contd The solid line shows $f(x)$ as a function of x while the dashed line shows the linear interpolation of f between $x^\star$ and $x^{\star\star}$.

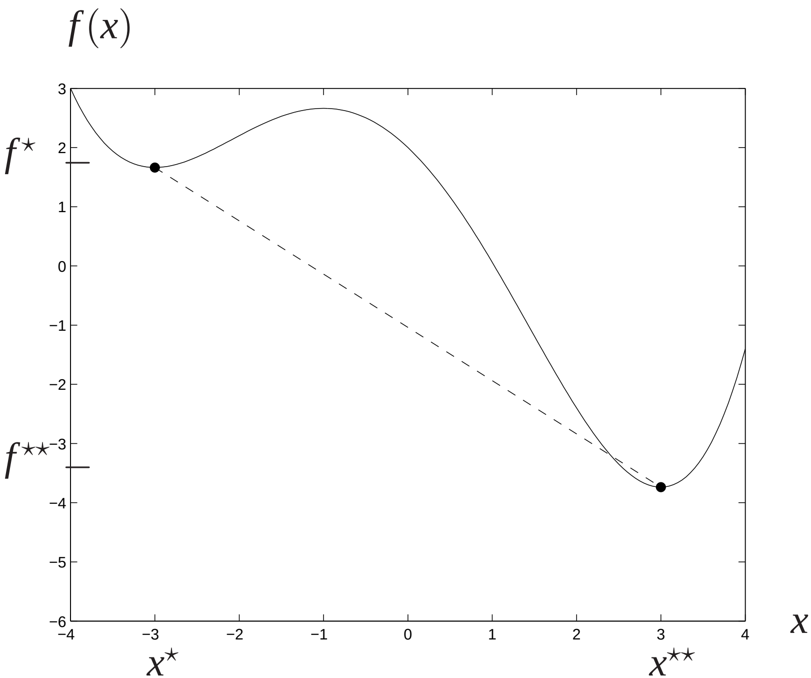


Fig. 2.32. Multiple minima and minimizers in proof of Theorem 2.4, Item (i).

Proof of (i), continued We are going to show that $x^\star$ satisfies the condition (2.27) for $x^\star$ not to be a local minimizer, which we repeat here for reference:

$$\forall \varepsilon > 0, \exists x^\varepsilon \in \mathbb{S} \text{ such that } (\|x^\star - x^\varepsilon\| < \varepsilon) \text{ and } (f(x^\star) > f(x^\varepsilon)).$$

We have:

$$\begin{aligned}
&\forall t \in [0,1], f(x^\star + t[x^{\star\star} - x^\star]) \\
&\quad \leq \quad f(x^\star) + t[f(x^{\star\star}) - f(x^\star)], \text{ by convexity of } f, \\
&\quad = \quad f^\star + t[f^{\star\star} - f^\star], \text{ by definition of } f^\star \text{ and } f^{\star\star}, \\
&\quad < \quad f^\star, \text{ for } 0 < t \leq 1, \text{ since } f^\star > f^{\star\star}, \\
&\quad = \quad f(x^\star).
\end{aligned} \tag{2.30}$$

Proof of (i), continued For $0 \leq t \leq 1$, we have $x^\star + t(x^{\star\star} - x^\star) \in \mathbb{S}$ since $\mathbb{S}$ is convex. But this means that there are feasible points arbitrarily close to $x^\star$ that have a lower objective value. In particular, given any norm $\|\bullet\|$ and any number $\varepsilon > 0$, we can define $x^\varepsilon = x^\star + t(x^{\star\star} - x^\star)$ where t is specified by:

$$t = \min\left\{1, \frac{\varepsilon}{2\|x^{\star\star} - x^\star\|}\right\}.$$

Proof of (i), continued Note that $x^\varepsilon \in \mathbb{S}$ since $0 \leq t \leq 1$ and that x^ε satisfies:

$$\begin{aligned}
\|x^\star - x^\varepsilon\| &= \|x^\star - [x^\star + t(x^{\star\star} - x^\star)]\|, \text{ by definition of } x^\varepsilon, \\
&= \|-t(x^{\star\star} - x^\star)\|, \\
&= |t| \times \|x^{\star\star} - x^\star\|, \text{ by a property of norms,} \\
&\leq \frac{\varepsilon}{2\|x^{\star\star} - x^\star\|}\|x^{\star\star} - x^\star\|, \text{ by definition of } t, \\
&= \frac{1}{2}\varepsilon, \\
&< \varepsilon.
\end{aligned}$$

Proof of (i), continued Furthermore $0 < t \leq 1$ by construction, so by (2.30):

$$f(x^\star) > f(x^\varepsilon).$$

That is, $x^\star$ satisfies (2.27) and is therefore not a local minimizer of f, which is a contradiction. As suggested by the "hump" in f at $x \approx -1$, the situation illustrated in Figure 2.32 is inconsistent with the assumption that f is convex. We conclude that f has at most one local minimum.

(ii) Suppose that the local minimum is $f^\star \in \mathbb{R}$ with corresponding local minimizer $x^\star \in \mathbb{S}$. Suppose that it is not a global minimum and minimizer. That is, there exists $x^{\star\star} \in \mathbb{S}$ such that $f^{\star\star} = f(x^{\star\star}) < f(x^\star)$. Then the same argument as in Item (i) shows that $f^\star$ is not a local minimum.

(iii) Suppose that f is strictly convex, yet that it has two local minimizers, $x^\star \neq x^{\star\star}$, say. Since f is convex, then by Item (i), both minimizers correspond to the unique minimum, say $f^\star$, of f over $\mathbb{S}$. We have:

$$\begin{aligned}
\forall t \in (0,1), f(x^\star + t[x^{\star\star} - x^\star]) &< f(x^\star) + t[f(x^{\star\star}) - f(x^\star)], \\
&\qquad \text{by strict convexity of } f, \\
&= f^\star + t[f^\star - f^\star], \text{ by definition of } f^\star, \\
&= f^\star,
\end{aligned}$$

which means that neither $x^\star$ nor $x^{\star\star}$ were local minimizers of f, since feasible points of the form $x^\star + t(x^{\star\star} - x^\star)$ have a lower objective value for all $t \in (0,1)$. That is, by a similar argument to that in Item (i), we can construct a feasible x^ε that is within a distance ε of $x^\star$ having a smaller value of objective than $x^\star$.
$\square$

Definition 2.19 If $\mathbb{S} \subseteq \mathbb{R}^n$ is a convex set and $f : \mathbb{R}^n \to \mathbb{R}$ is convex on $\mathbb{S}$, then $\min_{x\in\mathbb{S}} f(x)$ is called a **convex problem**. $\square$

2.6.3.4 *Discussion*

Local versus global minimizers

- Theorem 2.4 shows that a convex problem has at most one local minimum.
- If we find a local minimum for a convex problem, it is in fact the global minimum.

Choice of step directions

- Convexity enables us to relate the two goals of:

  (i) moving from the current iterate in a direction that decreases the objective while still maintaining feasibility, and

  (ii) moving from the current iterate towards the minimizer of the problem.

- If we have a convex problem, then these goals are not inconsistent.

Convex problems

- If:
  - the function $f : \mathbb{R}^n \to \mathbb{R}$ is convex,
  - the function $g : \mathbb{R}^n \to \mathbb{R}^m$ is affine, with $\forall x \in \mathbb{R}^n, g(x) = Ax - b$, and
  - the function $h : \mathbb{R}^n \to \mathbb{R}^r$ is convex,
- then:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, h(x) \leq \mathbf{0} \}$$

- is a convex problem.
- Some problems involving non-convex functions also specify convex problems.

Weakening the convexity assumption

- For example, a function with convex level sets has only one local minimum.

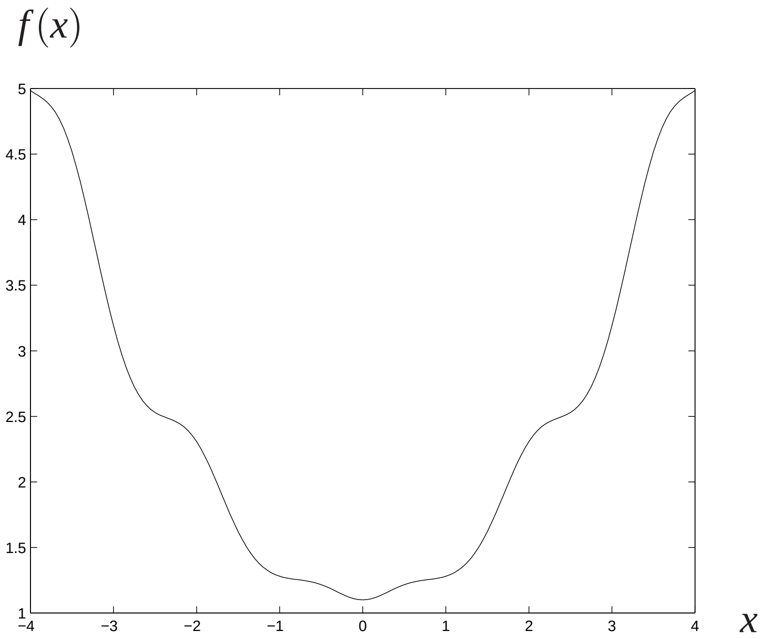


Fig. 2.33. A non-convex function with convex level sets.

Maximizing a convex function

Definition 2.20 Let $\mathbb{S} \subseteq \mathbb{R}^n$ and $x \in \mathbb{S}$. We say that x is an **extreme point** of $\mathbb{S}$ if:

$$\forall x', x'' \in \mathbb{S}, \left((x' \neq x) \text{ and } (x'' \neq x)\right) \Rightarrow \left(x \neq \frac{1}{2}(x' + x'')\right).$$

$\square$

- That is, x is an extreme point of $\mathbb{S}$ if it cannot be expressed as the "average" of two other points in $\mathbb{S}$.
- In Figure 2.22, there are three polygons:
  - The extreme points of each polygon are its vertices.
- The extreme points of the dodecahedron in Figure 2.14 are its vertices.
- In Figure 2.22, the extreme points of the filled ellipse are the points on the ellipse.

Theorem 2.5 *Let $\mathbb{S} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathbb{S} \to \mathbb{R}$ be convex on $\mathbb{S}$. Consider the maximization problem:*

$$\max_{x \in \mathbb{S}} f(x),$$

Suppose this problem possesses a maximum. Then there is a maximizer of this problem that is an extreme point of $\mathbb{S}$. $\square$

- In principle, we can maximize a convex objective over a convex set by searching over all the extreme points of the feasible set.
- There may be a very large number of extreme points of a set and this approach is not practical in general.
- However, for affine objectives and affine constraints (and some other cases), this approach leads to a practical method of optimization: the **simplex method** of linear programming.

We will discuss the simplex method in Chapter 16.

2.6.3.5 Characterizing convex functions

First partial derivatives

Theorem 2.6 *Let $\mathbb{S} \subseteq \mathbb{R}^n$ be a convex set and suppose that $f : \mathbb{S} \to \mathbb{R}$ is partially differentiable with continuous partial derivatives on $\mathbb{S}$. Then f is convex on $\mathbb{S}$ if and only if:*

$$\forall x, x' \in \mathbb{S}, f(x) \geq f(x') + \nabla f(x')^\dagger (x - x'). \tag{2.31}$$

$\square$

- The function $\phi : \mathbb{R}^n \to \mathbb{R}$ on the right-hand side of (2.31) defined by:

$$\forall x \in \mathbb{R}^n, \phi(x) = f(x') + \nabla f(x')^\dagger (x - x'),$$

- is called the **first-order Taylor approximation** of the function f, linearized about x'.

First-order Taylor expansion

- The inequality in (2.31) shows that the first-order Taylor approximation of a convex function never over-estimates the function.

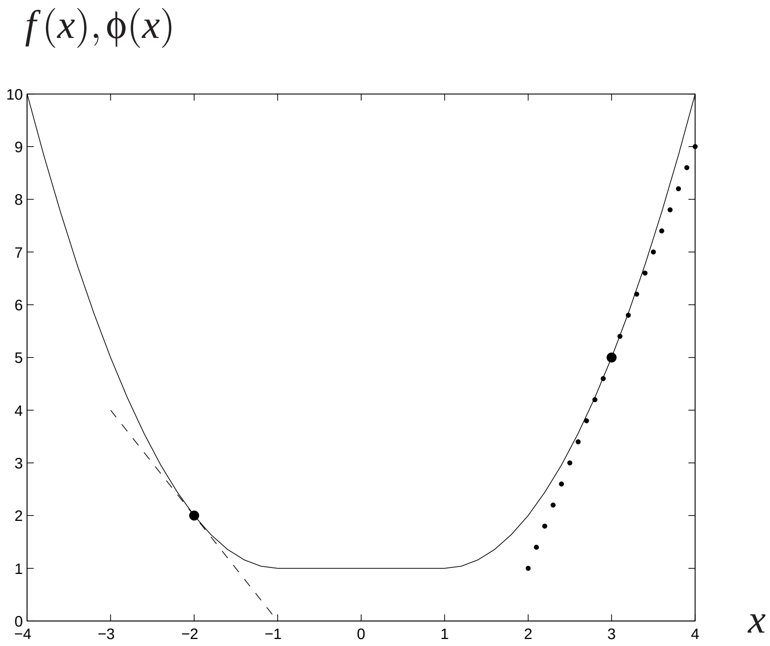


Fig. 2.34. First-order Taylor approximation about $x = -2$ (shown dashed) and about $x = 3$ (shown dotted) of a convex function (shown solid).

Sandwiching of convex function

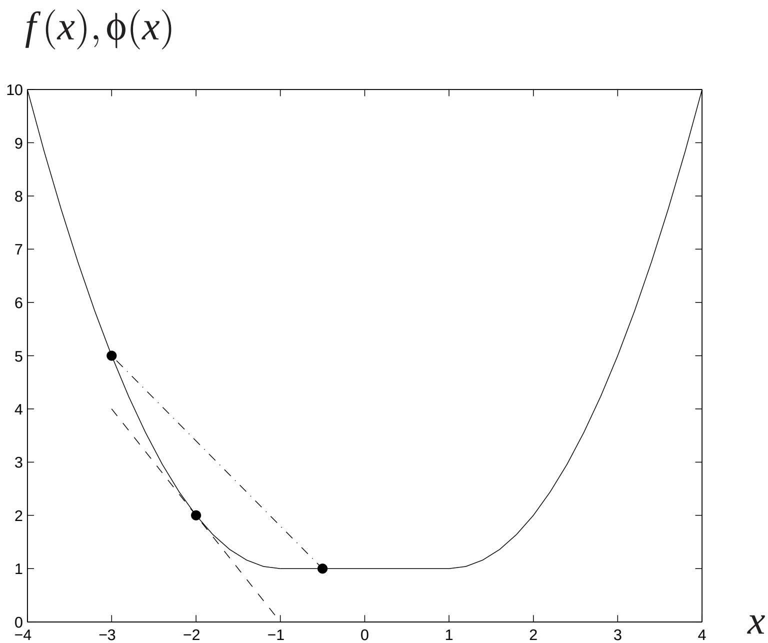


Fig. 2.35. Sandwiching of convex function between two affine functions. The first-order Taylor approximation about $x = -2$ (shown dashed) is a lower bound to the function. The linear interpolation of f between $x = -3$ and $x = -0.5$ (shown dash-dotted) is an upper bound to the function on the interval $\{x \in \mathbb{R} | -3 \leq x \leq -0.5\}$.

Second partial derivatives

- There are also tests of convexity involving positive semi-definiteness of the matrix of second partial derivatives, which is called the **Hessian** and is denoted $\nabla^2 f$ or $\nabla^2_{xx} f$.

Theorem 2.7 *Let $\mathbb{S} \subseteq \mathbb{R}^n$ be convex and suppose that $f : \mathbb{S} \to \mathbb{R}$ is twice partially differentiable with continuous second partial derivatives on $\mathbb{S}$. Suppose that the second derivative $\nabla^2 f$ is positive semi-definite throughout $\mathbb{S}$. Then f is convex on $\mathbb{S}$. If $\nabla^2 f$ is positive definite throughout $\mathbb{S}$ then f is strictly convex throughout $\mathbb{S}$.* $\square$

2.6.3.6 Further examples of convex functions

Quadratic functions

$$\forall x \in \mathbb{R}^n, f(x) = \frac{1}{2}x^\dagger Q x + c^\dagger x, \tag{2.32}$$

- where $Q \in \mathbb{R}^{n\times n}$ and $c \in \mathbb{R}^n$ are constants and Q is symmetric.
- The Hessian of this function is Q, which is constant and independent of x.
- If Q is positive semi-definite then, by Theorem 2.7, f is convex.
- If Q is positive definite then, by Theorem 2.7, f is strictly convex.

Piece-wise functions may or may not be convex

$$\forall x \in \mathbb{R}, f(x) = \begin{cases} (x+5)^2, & \text{if } x \leq 0, \\ (x-5)^2, & \text{if } x > 0. \end{cases}$$

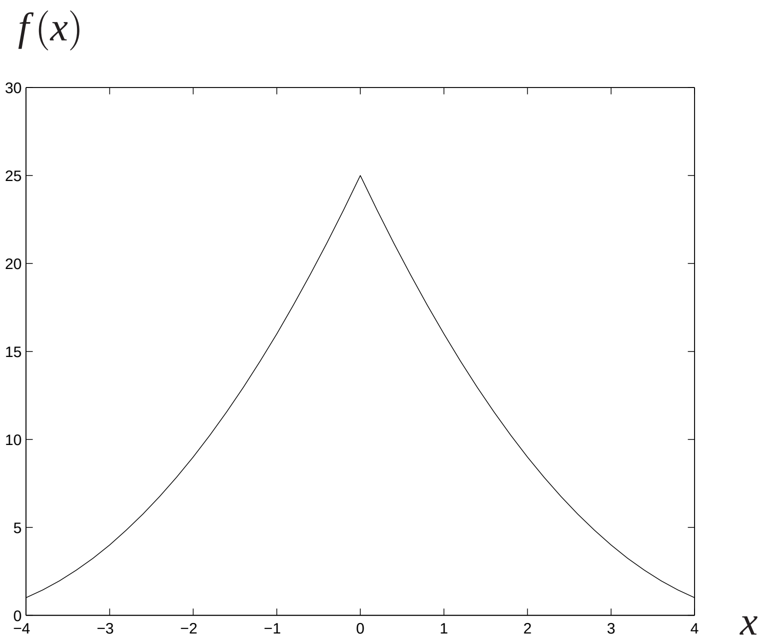


Fig. 2.36. Example of a piece-wise quadratic non-convex function.

Point-wise maxima of convex functions are convex

$$\forall x \in \mathbb{R}^n, f(x) = \max_{\ell=1,\ldots,r} f_\ell(x). \tag{2.33}$$

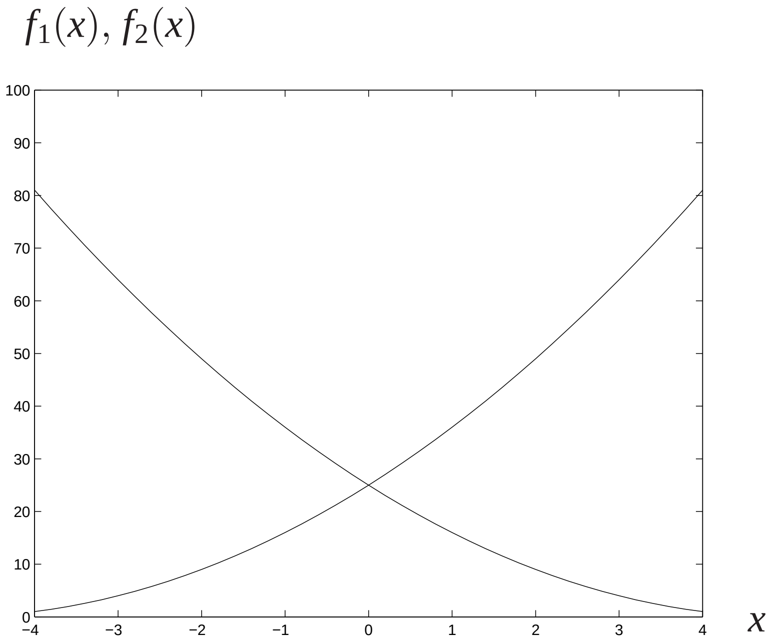


Fig. 2.37. Functions used to define point-wise maximum.

Point-wise maxima, continued

$$\forall x \in \mathbb{R},\ f(x) = \max\{f_1(x), f_2(x)\} = \begin{cases} (x+5)^2, & \text{if } x \geq 0, \\ (x-5)^2, & \text{if } x < 0. \end{cases}$$

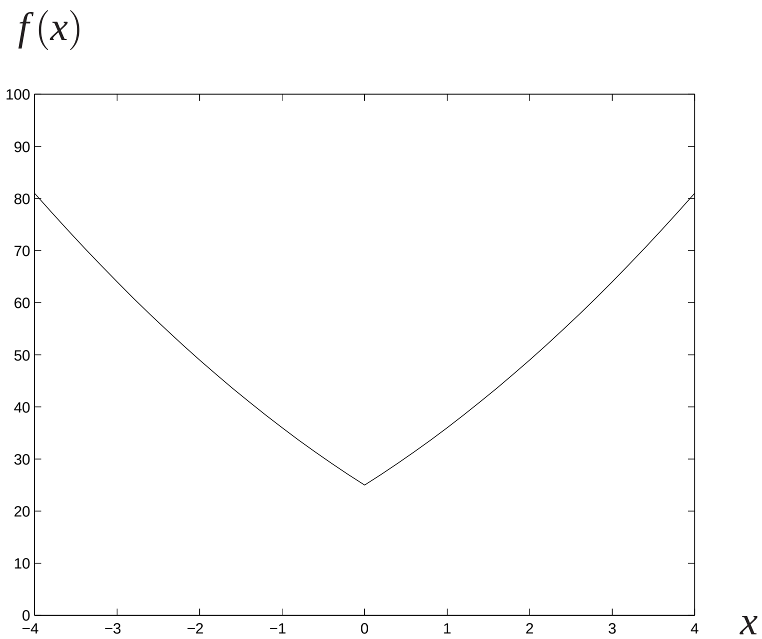


Fig. 2.38. Example of a piece-wise quadratic convex function.

2.7 Sensitivity and large change analysis

2.7.1 Motivation

- In many cases, the solution of a particular set of simultaneous equations or a particular optimization problem forms only a part of a larger design process in which the definition of the problem can be changed.

2.7.2 Parameterization

- Let us represent the change in the problem by supposing that the problem is *parameterized* by a vector $\chi \in \mathbb{R}^s$.
- For example, for linear equations:

$$A(\chi)x = b(\chi).$$

- We solve the base-case equations $A(\mathbf{0})x = b(\mathbf{0})$ for a base-case solution $x^\star$.
- We might then want to solve the equations for another value of χ and we consider solving $A(\chi)x = b(\chi)$ for the change-case solution.

Parameterization, continued

- Non-linear equations:

$$g(x;\chi) = \mathbf{0}.$$

- Optimization:

$$\min_{x \in \mathbb{R}^n} \{ f(x;\chi) | g(x;\chi) = \mathbf{0}, h(x;\chi) \leq \mathbf{0} \}.$$

2.7.3 *Sensitivity*

- We calculate the partial derivatives of the minimum and minimizer with respect to entries of χ, evaluated at the base-case solution corresponding to $\chi = \mathbf{0}$, and estimate the change in the solution based on the partial derivatives.
- Abusing notation, we will consider $f^\star$ and $x^\star$ to be *functions* of χ and write $\frac{\partial f^\star}{\partial \chi}$ and $\frac{\partial x^\star}{\partial \chi}$ for the sensitivities of the minimum and minimizer with respect to χ.
- We will generally only evaluate these sensitivities for $\chi = \mathbf{0}$.
- In general, we would prefer not to have to solve the change-case explicitly in order to calculate the derivatives.

2.7.4 *Large changes*

- By "large change" we mean a change that is so large that analysis based on the derivatives is or may be inaccurate.

2.7.5 Examples

- In this section we consider examples of sensitivity analysis for each of the five problem classes.

2.7.5.1 Linear simultaneous equations

$$\forall \chi \in \mathbb{R}, A(\chi) = \begin{bmatrix} 1 & 2+\chi \\ 3 & 4 \end{bmatrix}, b(\chi) = \begin{bmatrix} 1 \\ 1+\chi \end{bmatrix}, \tag{2.34}$$

2.7.5.2 Non-linear simultaneous equations

$$\forall x \in \mathbb{R}, \forall \chi \in \mathbb{R}, g(x;\chi) = (x - 2 - \sin\chi)^3 + 1.$$

2.7.5.3 Unconstrained minimization

$$\forall x \in \mathbb{R}^2, \forall \chi \in \mathbb{R}, f(x;\chi) = (x_1 - \exp(\chi))^2 + (x_2 - 3\exp(\chi))^2 + 5\chi.$$

2.7.5.4 Equality-constrained minimization

$$\min_{x \in \mathbb{R}^2} \{ f(x) | Ax = b(\chi) \},$$

$$\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\
A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\
\forall \chi \in \mathbb{R}, b(\chi) &= [-\chi].
\end{aligned}$$

2.7.5.5 Inequality-constrained minimization

$$\min_{x \in \mathbb{R}^2} \{ f(x) | g(x) = 0, h(x;\chi) \leq 0 \},$$

$$\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\
\forall x \in \mathbb{R}^2, g(x) &= x_1 - x_2, \\
\forall x \in \mathbb{R}^2, \forall \chi \in \mathbb{R}, h(x;\chi) &= 3 - x_2 - \chi.
\end{aligned}$$

2.7.6 *Ill-conditioned problems*

2.7.6.1 *Motivation*

Definition 2.21 A problem is said to be **ill-conditioned** if a relatively small change in the problem specification leads to a relatively large change in the solution. □

2.7.6.2 *Simultaneous equations example*

- Consider simultaneous equations that are redundant.
- For example, suppose that two entries, g_1 and g_2, of $g : \mathbb{R}^n \to \mathbb{R}^m$ are the same.
- Suppose that $x^\star$ is a solution of $g(x) = \mathbf{0}$, so that $g_1(x^\star) = g_2(x^\star) = 0$.
- An arbitrarily small change in the problem specification results in a large qualitative change in the solution: the problem can change from having a solution to having no solution.
- That is, redundant simultaneous equations are ill-conditioned.
- For this reason, we will generally try to avoid redundant equations in the formulation of simultaneous equations problems and avoid redundancy in formulating equality constraints in optimization problems.

2.7.6.3 Optimization example

- Suppose that we wish to minimize a convex function and consider the problem of finding a step direction that points towards the minimizer of the problem based on "local" first derivative information about the function at a particular iterate $x^{(\nu)}$.
- The direction *perpendicular* to the surface of the contour set at a point is particularly easy to find.
- This direction is the negative of the gradient of f evaluated at the point.

Example

- For circular contour sets, the direction perpendicular to the surface of the contour set points directly towards the unconstrained minimizer of f.

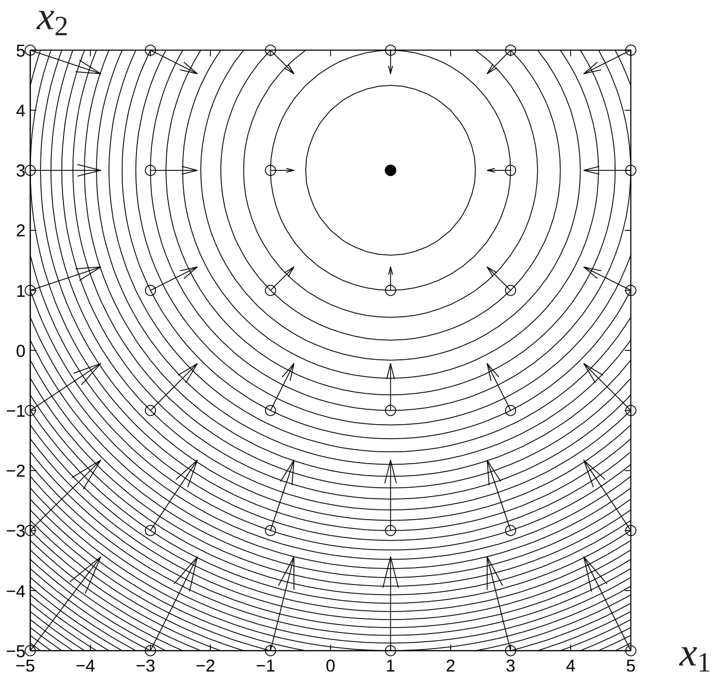


Fig. 2.39. Directions perpendicular to contour sets.

Example

- For elliptical contour sets, movement perpendicular to the contour set will not point directly towards the minimizer.

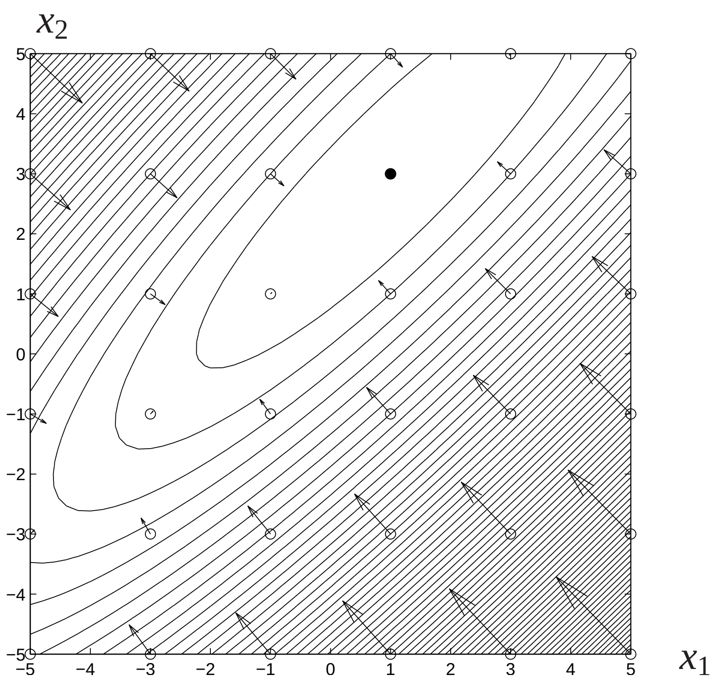


Fig. 2.40. Directions perpendicular to contour sets.

Eccentric contour sets

- If the contour sets are highly eccentric then the problem of using the gradient to find the direction that points towards the minimizer is ill-conditioned.
- Suppose that the function changes slightly, so that its minimizer is at $x^{\star\star} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ instead of $x^{\star} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.
- A contour plot of the changed function is shown in Figure 2.41.
- The arrows in Figure 2.41 are in essentially the same direction as those shown in Figure 2.40.
- The change in minimizer has had negligible effect on the information provided by the direction perpendicular to the contour sets.
- The problem of finding a direction that points towards the minimizer using the information provided by the direction that is perpendicular to the contour set is ill-conditioned.

Eccentric contour sets, continued

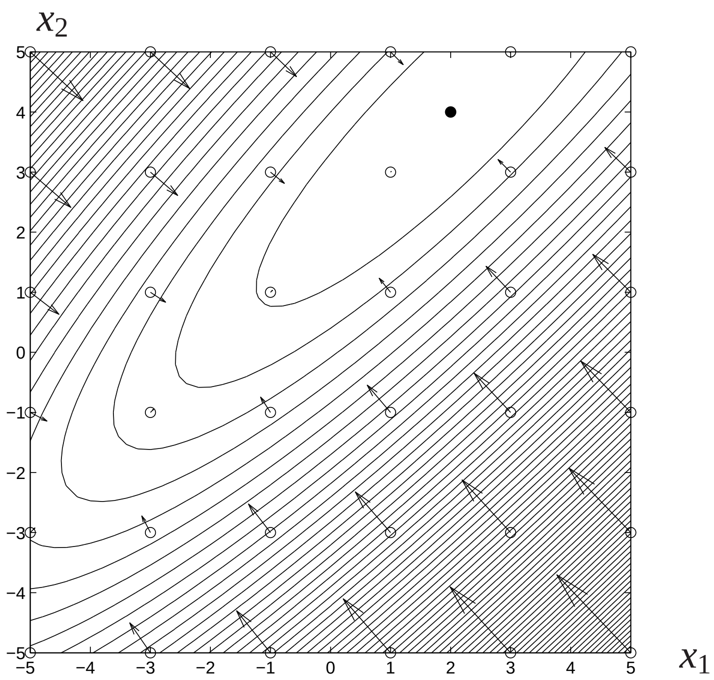


Fig. 2.41. Directions perpendicular to contour sets for changed function.

2.7.6.4 Discussion

- In both examples, small changes in the problem led to large changes in the solution, in either a qualitative or quantitative sense.
- We will consider ill-conditioning in several contexts throughout the course.

2.8 Summary

- In this chapter we have defined two main classes of problems:

  (i) simultaneous equations, and
  (ii) optimization problems,

- illustrating particular types of problems with elementary examples.
- We defined **direct** and **iterative** algorithms and characterized:
  - conditions for uniqueness of solution of simultaneous equations using the notion of a monotone function,
  - local and global and strict and non-strict minima and minimizers of optimization problems using the notion of convexity,
  - conditions for uniqueness of a local minimum and minimizer.
- We also discussed sensitivity analysis and ill-conditioned problems.

3

Transformation of problems

Outline

- Transformations of the objective in Section 3.1;
- Transformations of the variables in Section 3.2;
- Transformations of the constraints in Section 3.3; and
- Transformation of the problem involving a notion called "duality" in Section 3.4.

3.1 Objective

- Basic techniques for transforming the objective that we will discuss are:

  (i) **monotonically increasing transformations**,
  (ii) **adding terms**,
  (iii) **moving the objective into the constraints**, and
  (iv) **approximating the objective**.

3.1.1 Monotonically increasing transformations

Theorem 3.1 *Let $\mathbb{S} \subseteq \mathbb{R}^n$, let $f : \mathbb{R}^n \to \mathbb{R}$, and let $\eta^{\nearrow} : \mathbb{R} \to \mathbb{R}$ be strictly monotonically increasing on $\mathbb{R}$. Define $\phi : \mathbb{R}^n \to \mathbb{R}$ by:*

$$\forall x \in \mathbb{R}^n, \phi(x) = \eta^{\nearrow}(f(x)).$$

Consider the problems: $\min_{x\in\mathbb{S}} \phi(x)$ *and* $\min_{x\in\mathbb{S}} f(x)$. *Then:*

(i) $\min_{x\in\mathbb{S}} f(x)$ *has a minimum if and only if* $\min_{x\in\mathbb{S}} \phi(x)$ *has a minimum.*

(ii) *If either one of the problems in Item* (i) *possesses a minimum (and consequently, by Item* (i)*, each one possesses a minimum), then:*

$$\begin{aligned} \eta^{\nearrow}\left(\min_{x\in\mathbb{S}} f(x)\right) &= \min_{x\in\mathbb{S}} \phi(x), \\ \operatorname*{argmin}_{x\in\mathbb{S}} f(x) &= \operatorname*{argmin}_{x\in\mathbb{S}} \phi(x). \end{aligned}$$

$\square$

Discussion

- Two transformations of objective that will prove particularly useful in our case studies involve the exponential function and (for strictly positive objective) the logarithmic function.
- The squared function provides another example of a monotonically increasing transformation for a function $f : \mathbb{R}^n \to \mathbb{R}_+$.

3.1.2 Adding terms

- Consider adding terms that depend on the constraint function with a view to incorporating the constraints into the objective so that either:
  - we do not have to consider the constraints explicitly, or
  - the constraints are easier to deal with.

3.1.2.1 Penalty function

Theorem 3.2 *Let* $\mathbb{S} \subseteq \mathbb{R}^n$ *and* $f : \mathbb{R}^n \to \mathbb{R}$. *Consider the optimization problem* $\min_{x\in\mathbb{S}} f(x)$. *Let* $f_{\mathrm{p}} : \mathbb{R}^n \to \mathbb{R}_+$ *be such that* $(x \in \mathbb{S}) \Rightarrow (f_{\mathrm{p}}(x) = 0)$ *and let* $\Pi \in \mathbb{R}_+$. *Then:*

(i) $\min_{x\in\mathbb{S}} f(x)$ *has a minimum if and only if* $\min_{x\in\mathbb{S}}(f(x) + \Pi f_{\mathrm{p}}(x))$ *has a minimum.*

(ii) *If either one of the problems in Item* (i) *possesses a minimum (and consequently, by Item* (i)*, each one possesses a minimum), then:*

$$\begin{aligned} \min_{x\in\mathbb{S}} f(x) &= \min_{x\in\mathbb{S}}(f(x) + \Pi f_{\mathrm{p}}(x)), \\ \operatorname*{argmin}_{x\in\mathbb{S}} f(x) &= \operatorname*{argmin}_{x\in\mathbb{S}}(f(x) + \Pi f_{\mathrm{p}}(x)). \end{aligned}$$

□

Discontinuous penalty function

Example

- Consider the objective $f : \mathbb{R} \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}, f(x) = x. \tag{3.1}$$

- The problem:

$$\min_{x \in \mathbb{R}} \{ f(x) | 1 \leq x \leq 3 \},$$

- has minimum $f^\star = 1$ and minimizer $x^\star = 1$.
- Let $\Pi = 1$ and consider the penalty function $f_\mathrm{p} : \mathbb{R} \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}, f_\mathrm{p}(x) = \begin{cases} 0, & \text{if } 1 \leq x \leq 3, \\ 10, & \text{otherwise.} \end{cases} \tag{3.2}$$

Example, continued

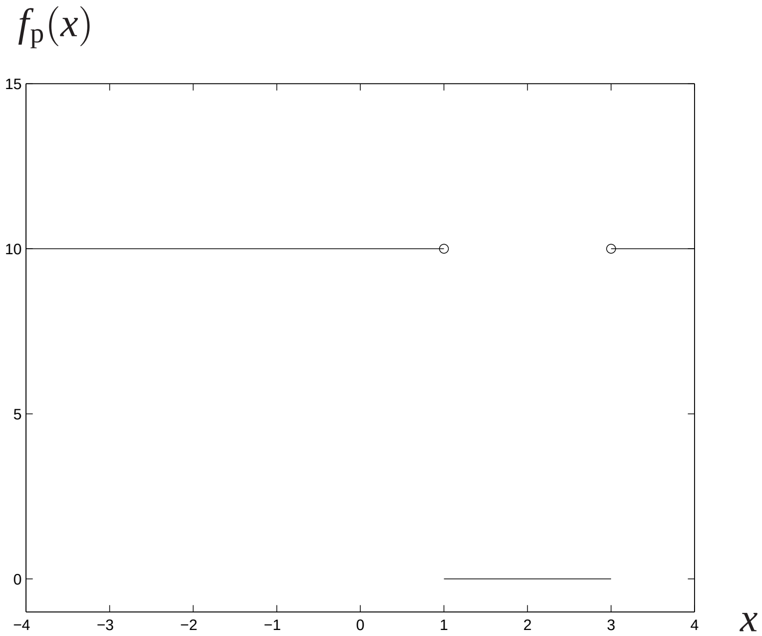


Fig. 3.1. The penalty function $f_p(x)$ versus x. In this figure and the next, the circles $\circ$ indicate that the function has a point of discontinuity as x approaches 1 from below or 3 from above.

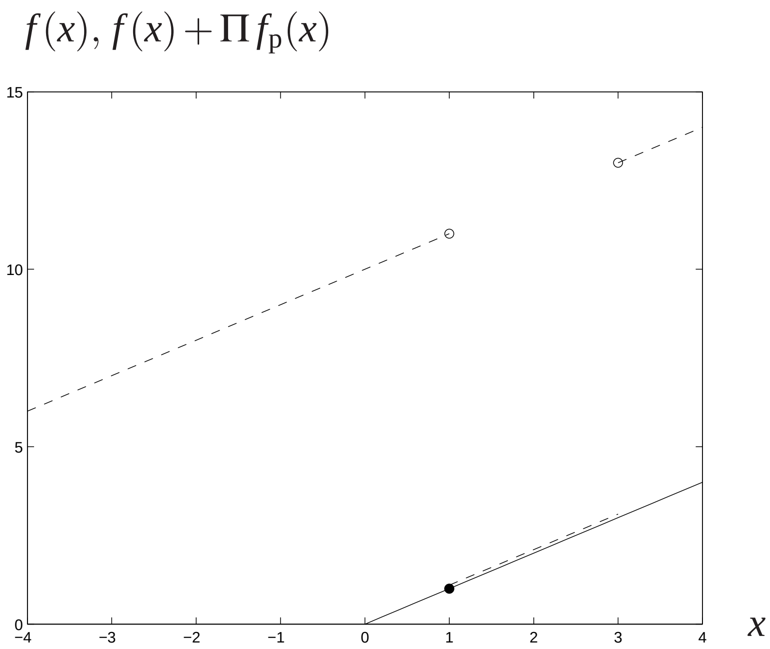


Fig. 3.2. The objective function $f(x)$ versus x (shown solid) and the penalized objective function $f(x) + \Pi f_p(x)$ versus x (shown dashed). (For clarity, for $1 \leq x \leq 3$ the penalized objective function is drawn slightly above the solid curve: it should be coincident with the solid curve.) One local minimizer of $f + \Pi f_p$ in the region $\{x \in \mathbb{R} | -4 \leq x \leq 4\}$ is indicated by the bullet $\bullet$.

Example, continued

- The point $x^\star = 1$ is an unconstrained local minimizer of $f + \Pi f_{\text{p}}$ in the region $\{x \in \mathbb{R} | -4 \leq x \leq 4\}$ and is indicated in Figure 3.2 by a bullet $\bullet$.
- The penalty function allows us to consider the effect of the constraints by considering the penalized objective only.

Discussion

- The drawback of the penalty function f_{p} defined in (3.2) is that the penalized objective function $f + \Pi f_{\text{p}}$ is not continuous because of the form of f_{p}.
- Moreover, local information at a feasible point in the interior $\underline{\mathbb{S}} = \{x \in \mathbb{R} | 1 < x < 3\}$ of $\mathbb{S}$ does not inform about the boundary of the feasible region.

Continuous penalty function

Corollary 3.3 *Suppose that* $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^m, \Pi \in \mathbb{R}_+$, *and that* $\|\bullet\|$ *is a norm on* $\mathbb{R}^m$. *Consider the optimization problems* $\min_{x\in\mathbb{R}^n}\{f(x)|g(x)=\mathbf{0}\}$ *and* $\min_{x\in\mathbb{R}^n}\{f(x)+\Pi\|g(x)\|^2|g(x)=\mathbf{0}\}$. *Then:*

(i) *the problem* $\min_{x\in\mathbb{R}^n}\{f(x)|g(x)=\mathbf{0}\}$ *has a minimum if and only if the problem* $\min_{x\in\mathbb{R}^n}\{f(x)+\Pi\|g(x)\|^2|g(x)=\mathbf{0}\}$ *has a minimum.*

(ii) *If either one of the problems in Item* (i) *possesses a minimum (and consequently, by Item* (i), *each one possesses a minimum), then*

$$\begin{aligned}\min_{x\in\mathbb{R}^n}\{f(x)|g(x)=\mathbf{0}\} &= \min_{x\in\mathbb{R}^n}\{f(x)+\Pi\|g(x)\|^2|g(x)=\mathbf{0}\},\\ \arg\min_{x\in\mathbb{R}^n}\{f(x)|g(x)=\mathbf{0}\} &= \arg\min_{x\in\mathbb{R}^n}\{f(x)+\Pi\|g(x)\|^2|g(x)=\mathbf{0}\}.\end{aligned}$$

Proof In the hypothesis of Theorem 3.2, let $\mathbb{S} = \{x \in \mathbb{R}^n | g(x) = \mathbf{0}\}$ and define $f_\mathrm{p} : \mathbb{R}^n \to \mathbb{R}_+$ by $f_\mathrm{p}(\bullet) = \|g(\bullet)\|^2$. Then:

$$\begin{aligned}(x \in \mathbb{S}) &\Leftrightarrow (g(x) = \mathbf{0}),\\ &\Rightarrow (f_\mathrm{p}(x) = 0),\end{aligned}$$

so that the hypothesis and therefore the conclusion of Theorem 3.2 holds. □

Example

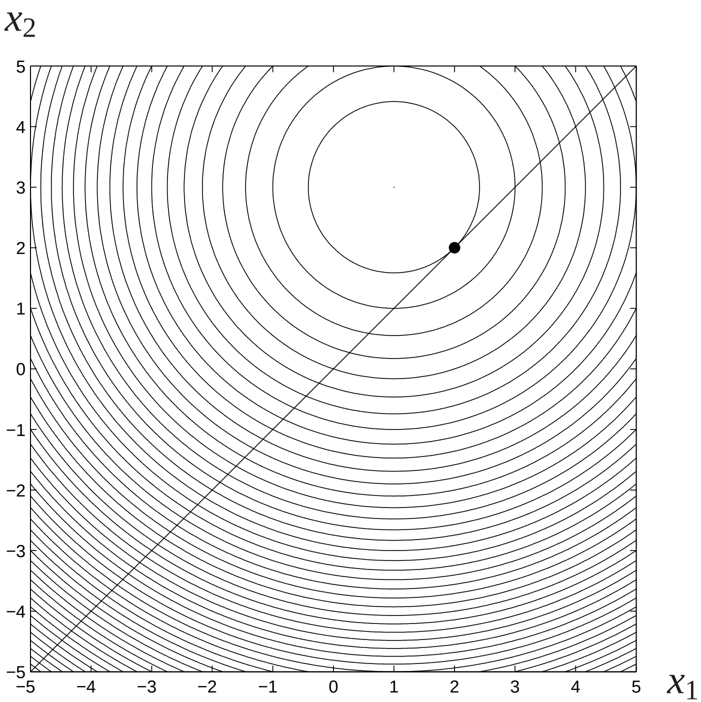


Fig. 3.3. Contour sets $\mathbb{C}_f(\tilde{f})$ of objective function and the feasible set from Problem (2.13). The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

Example, continued

- Figure 3.3 shows the contour sets of the objective $f : \mathbb{R}^2 \to \mathbb{R}$ of Problem (2.13):

$$\forall x \in \mathbb{R}^2, f(x) = (x_1 - 1)^2 + (x_2 - 3)^2,$$

- and a line that represents the feasible set $\{x \in \mathbb{R}^2 | g(x) = 0\}$, where $g : \mathbb{R}^2 \to \mathbb{R}$ is defined by:

$$\forall x \in \mathbb{R}^2, g(x) = x_1 - x_2.$$

- As discussed in Section 2.2.2, the minimizer of $\min_{x \in \mathbb{R}^2}\{f(x) | g(x) = 0\}$ is $x^\star = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.
- Figure 3.4 shows the contour sets of $(g(\bullet))^2$.
- The contours are parallel, since g is affine, and decrease towards the line representing the feasible set.

Example, continued

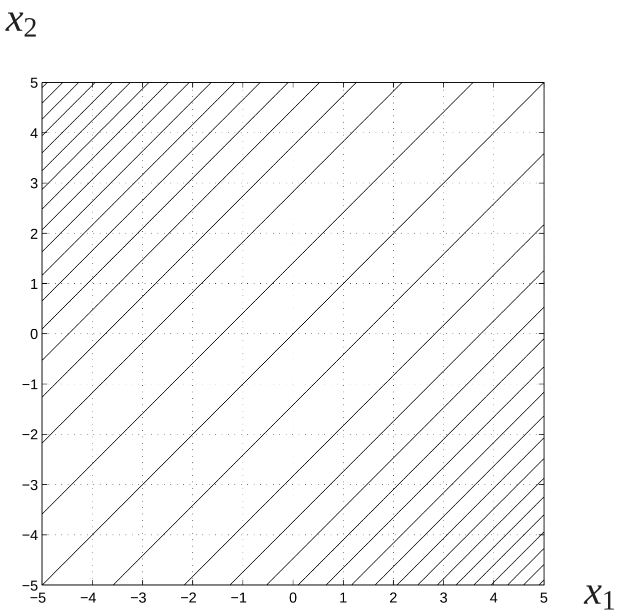


Fig. 3.4. Contour sets $\mathbb{C}_{(g)^2}(\tilde{f})$ of $(g(\bullet))^2$. The heights of the contours decrease towards the line $x_1 = x_2$.

Example, continued

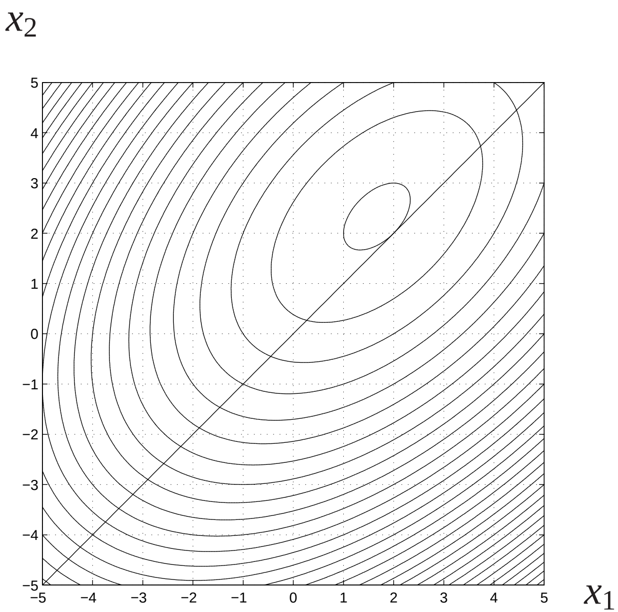


Fig. 3.5. Contour sets $\mathbb{C}_{f+1(g)^2}(\tilde{f})$ of penalized objective function and the feasible set from Problem (2.13). The heights of the contours decrease towards the point $\begin{bmatrix} 5/3 \\ 7/3 \end{bmatrix}$.

Example, continued

- Figure 3.5 shows the contour sets of the corresponding penalized objective $f(\bullet) + \Pi(g(\bullet))^2$ for $\Pi = 1$, and again shows the line representing the feasible set.
- Adding the penalty to the objective makes infeasible points less "attractive" and does not change the objective values on the feasible set.
- The unconstrained minimizer of $f(\bullet) + \Pi(g(\bullet))^2$ for $\Pi = 1$ is $\begin{bmatrix} 5/3 \\ 7/3 \end{bmatrix}$, which is closer to the minimizer of the equality-constrained problem than is the unconstrained minimizer of f.

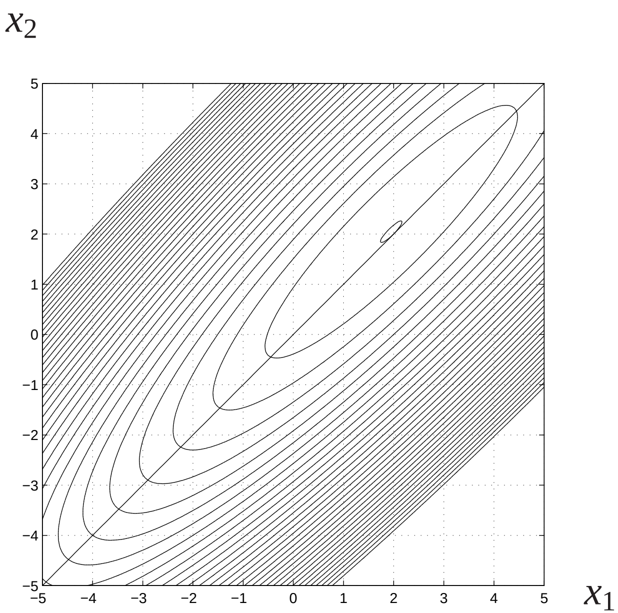


Fig. 3.6. Contour sets $\mathbb{C}_{f+10(g)^2}(\tilde{f})$ of penalized objective function and the feasible set from Problem (2.13). The heights of the contours decrease towards a point that is near to $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

Example, continued

- Larger values of the penalty coefficient Π, such as $\Pi = 10$ as shown in Figure 3.6, make infeasible points even less attractive.
- The unconstrained minimizer of $f(\bullet) + \Pi(g(\bullet))^2$ for $\Pi = 10$ is very close to $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$, which is the minimizer of the equality-constrained problem.

Sequence of problems

- Under certain conditions, the sequence of solutions of *unconstrained* problems approaches a solution of the constrained problem as $\Pi \to \infty$.

Soft constraints

- A penalty approach can be a very effective means to approximately satisfy "soft constraints."

Ill-conditioning

- For very tight tolerances, the required value of Π will be large.

- As Π becomes large the unconstrained problem becomes difficult to solve.

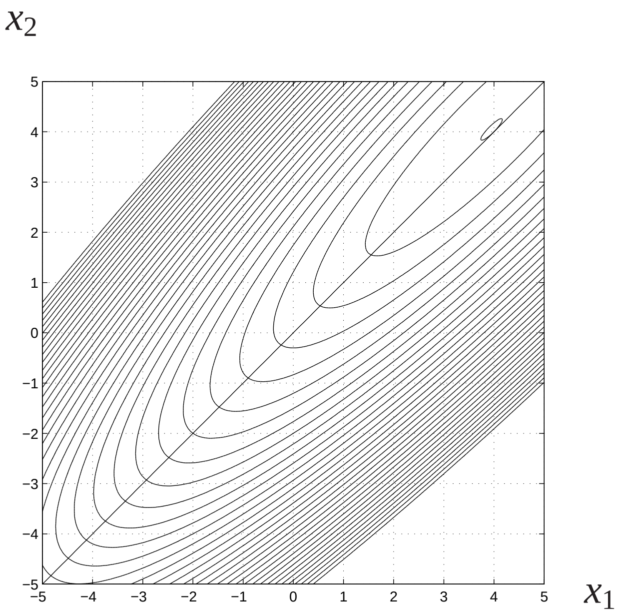


Fig. 3.7. The contour sets from Figure 3.6 shifted up and to the right. The feasible set from Problem (2.13) is also shown. The heights of the contours decrease towards a point that is near to $\begin{bmatrix} 4 \\ 4 \end{bmatrix}$.

Ill-conditioning, continued

- Figure 3.7 shows the case where the center of the ellipses are shifted up by two units and to the right by two units.
- As in the example in Section 2.7.6.3, the effect on local appearance of the level sets at a point such as $x = \begin{bmatrix} 0 \\ -5 \end{bmatrix}$ is only small.
- If $x^{(\nu)} = \begin{bmatrix} 0 \\ -5 \end{bmatrix}$ were the current iterate, for example, then it would be difficult to accurately determine the direction of the minimizer from local first derivative information at this point.

3.1.2.2 Barrier function

- Another approach to enforcing inequality constraints involves adding a function that grows large as we *approach* the boundary of the feasible region from the interior.
- Consider again the feasible set $\mathbb{S} = \{x \in \mathbb{R} | 1 \leq x \leq 3\}$ and its interior, $\underline{\mathbb{S}} = \{x \in \mathbb{R} | 1 < x < 3\}$.

- Figure 3.8 shows a function $f_{\mathrm{b}} : \underline{\mathbb{S}} \rightarrow \mathbb{R}$ that is designed to penalize values of x that are close to the boundary of the feasible region.

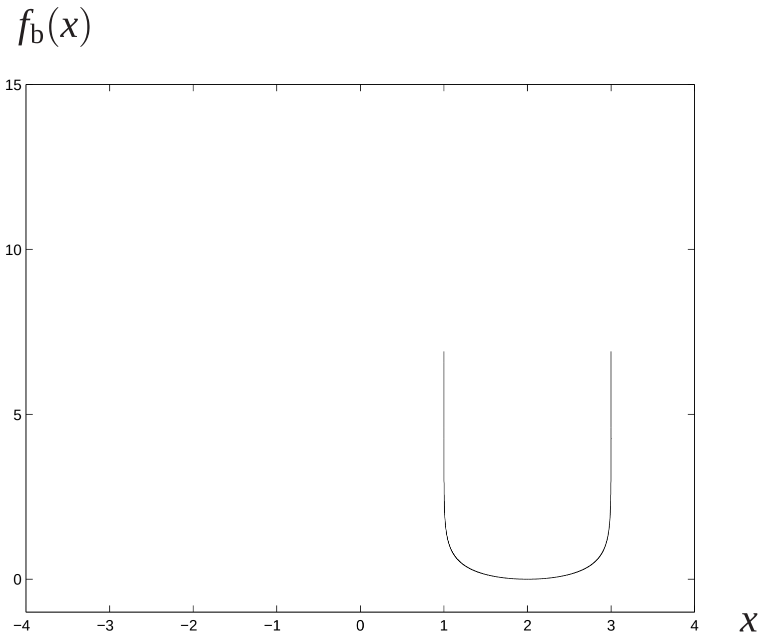


Fig. 3.8. The barrier function $f_{\mathrm{b}}(x)$ versus x on the interior of the feasible set.

Barrier function

- Consider again the objective function $f : \mathbb{R} \to \mathbb{R}$ defined in (3.1) and illustrated in Figure 3.2.
- Figure 3.9 shows this objective together with $f(x) + f_{\mathrm{b}}(x)$ for values of x that are in the interior of the feasible set $\{x \in \mathbb{R} | 1 \leq x \leq 3\}$.
- A local minimizer of $f + f_{\mathrm{b}}$ is illustrated with a $\bullet$.
- This point is nearby to the minimizer of the original constrained problem $\min_{x \in \mathbb{R}} \{f(x) | x \in \mathbb{S}\}$.
- We solve a *sequence* of problems where the added term is gradually reduced towards zero.

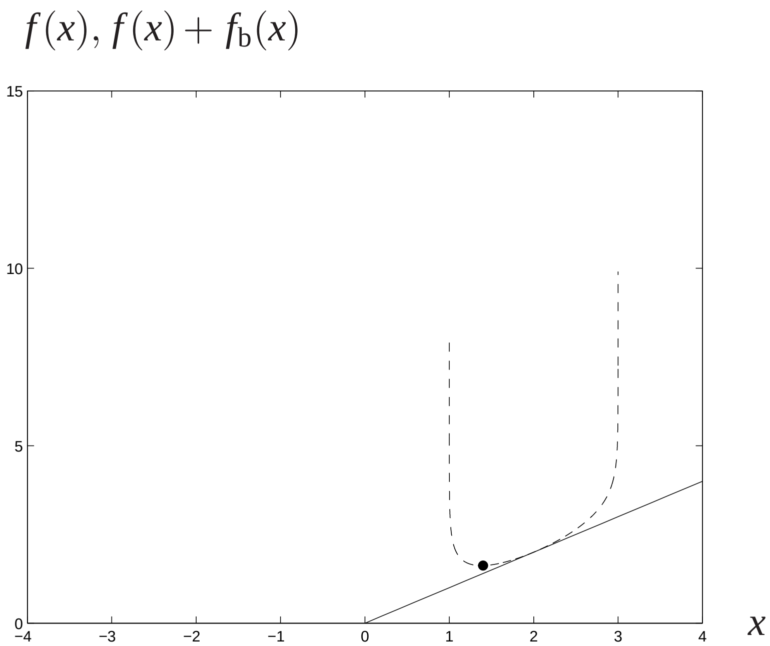


Fig. 3.9. The objective function $f(x)$ versus x (shown solid) and the objective plus barrier function $f(x) + f_b(x)$ versus x on the interior of the feasible set (shown dashed). The local minimizer of the objective plus barrier function is indicated by the bullet •.

3.1.3 Moving the objective into the constraints

$$\forall x \in \mathbb{R}, f(x) = \max\{(x+5)^2, (x-5)^2\}.$$

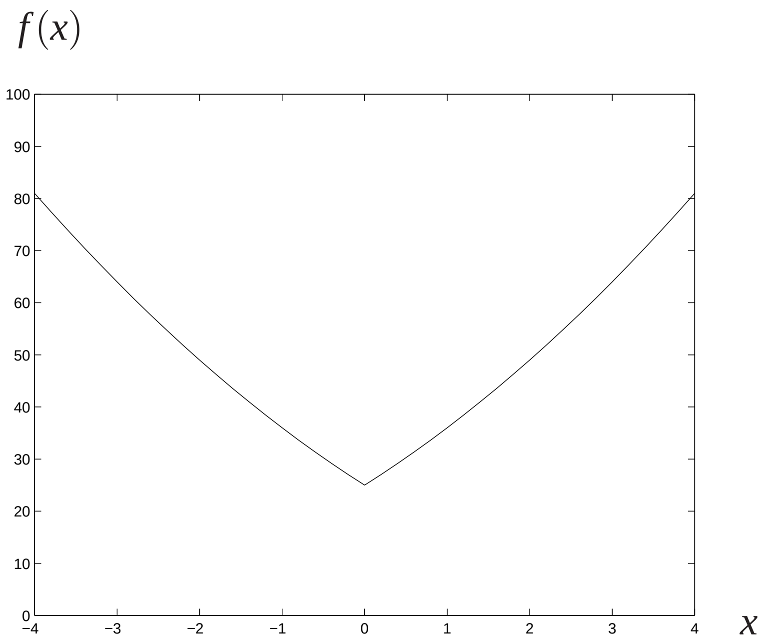


Fig. 3.10. Function defined as point-wise maximum.

Theorem 3.4 *Let $\mathbb{S} \subseteq \mathbb{R}^n$ and let $f_\ell : \mathbb{R}^n \to \mathbb{R}$ for $\ell = 1, \ldots, r$. Define $f : \mathbb{R}^n \to \mathbb{R}$ by:*

$$\forall x \in \mathbb{R}^n, f(x) = \max_{\ell=1,\ldots,r} f_\ell(x).$$

Consider the problems $\min_{x\in\mathbb{S}} f(x)$ and

$$\min_{x\in\mathbb{S}, z\in\mathbb{R}} \{z | f_\ell(x) - z \leq 0, \forall \ell = 1, \ldots, r\}. \tag{3.3}$$

Then:

(i) *the problem $\min_{x\in\mathbb{S}} f(x)$ has a minimum if and only if $\min_{x\in\mathbb{S},z\in\mathbb{R}}\{z | f_\ell(x) - z \leq 0, \forall \ell = 1, \ldots, r\}$ has a minimum.*

(ii) *If either one of the problems in Item (i) possesses a minimum (and consequently, by Item (i), each one possesses a minimum), then*

$$\begin{aligned}
\min_{x\in\mathbb{S}} f(x) &= \min_{x\in\mathbb{S},z\in\mathbb{R}} \{z | f_\ell(x) - z \leq 0, \forall \ell = 1, \ldots, r\}, \\
\operatorname*{argmin}_{x\in\mathbb{S}} f(x) &= \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{bmatrix} x \\ z \end{bmatrix} \in \arg\min_{x\in\mathbb{S},z\in\mathbb{R}} \left\{ z \,\middle|\, \begin{array}{l} f_\ell(x) - z \leq 0, \\ \forall \ell = 1, \ldots, r \end{array} \right\} \right\}.
\end{aligned}$$

□

Discussion

- Figure 3.11 repeats Figure 2.37 and shows the functions f_1 and f_2 that were point-wise maximized to form the objective shown in Figure 3.10.

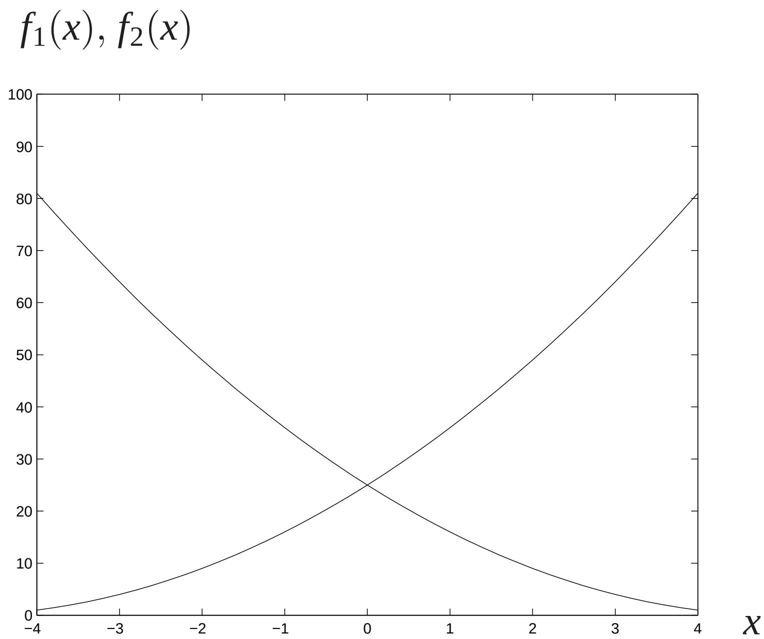


Fig. 3.11. The functions used to define point-wise maximum, repeated from Figure 2.37.

Discussion, continued

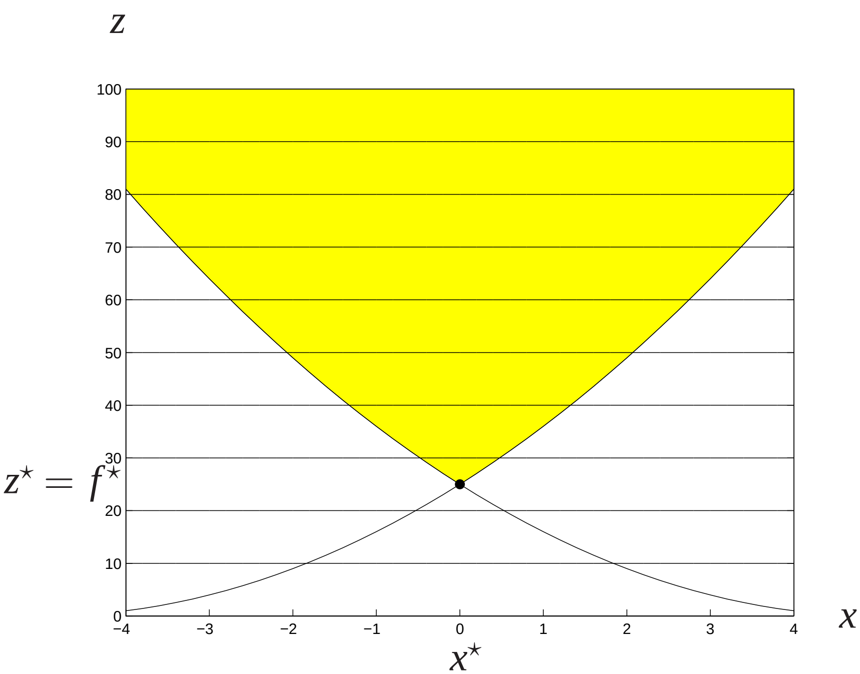


Fig. 3.12. Feasible region, shown shaded, and contour sets of objective for transformed problem. The feasible region is the set of points $\begin{bmatrix} x \\ z \end{bmatrix}$ that lies "above" both of the curves. The contour sets of the objective decrease towards $z = 0$.

Discussion, continued

- Figure 3.12 re-interprets Figure 3.11 in terms of Problem (3.3).
- It shows the feasible region, shown shaded, and the contour sets of the objective, which are lines of constant value of z.
- Problem (3.3) tries to find the minimum feasible value of z; that is, it seeks the "lowest" feasible line.

3.1.4 *Approximating the objective*

- The four basic techniques we will discuss are:

  (i) **linear approximation**,
  (ii) **quadratic approximation**,
  (iii) **piece-wise linearization**, and
  (iv) **smoothing**.

3.1.4.1 Linear approximation

- We linearize an objective about a current estimate $x^{(\nu)}$.
- A linear programming algorithm is then used to solve for the optimal $x^{(\nu+1)}$ that minimizes the linearized objective while satisfying the (linearized) constraints.
- Extra constraints are added to ensure that linear approximation is valid at the updated point:

$$\forall k = 1, \ldots, n, |x_k^{(\nu+1)} - x_k^{(\nu)}| \leq \overline{\Delta x_k}.$$

3.1.4.2 Quadratic approximation

- Instead of a linear approximation, a quadratic approximation can be made to the objective at each iteration ν.

3.1.4.3 *Piece-wise linearization*

- For a function $f : [0,1] \rightarrow \mathbb{R}$ we might:
  - define subsidiary variables $\xi_1, \ldots, \xi_5$,
  - include constraints:

$$\begin{aligned} x &= \sum_{j=1}^{5} \xi_j, \\ 0 &\leq \xi_j \leq 0.2, \end{aligned}$$

  - define parameters:

$$\begin{aligned} d &= f(0), \\ c_j &= \frac{1}{0.2}\left[f(0.2 \times j) - f(0.2 \times (j-1))\right], j = 1, \ldots, 5, \end{aligned}$$

    and
  - replace the objective f by the piece-wise linearized objective $\phi : \mathbb{R}^5 \rightarrow \mathbb{R}$ defined by:

$$\forall \xi \in \mathbb{R}^5, \phi(\xi) = c^\dagger \xi + d.$$

Quadratic example function

$$\forall x \in [0,1],\, f(x) = (x)^2.$$

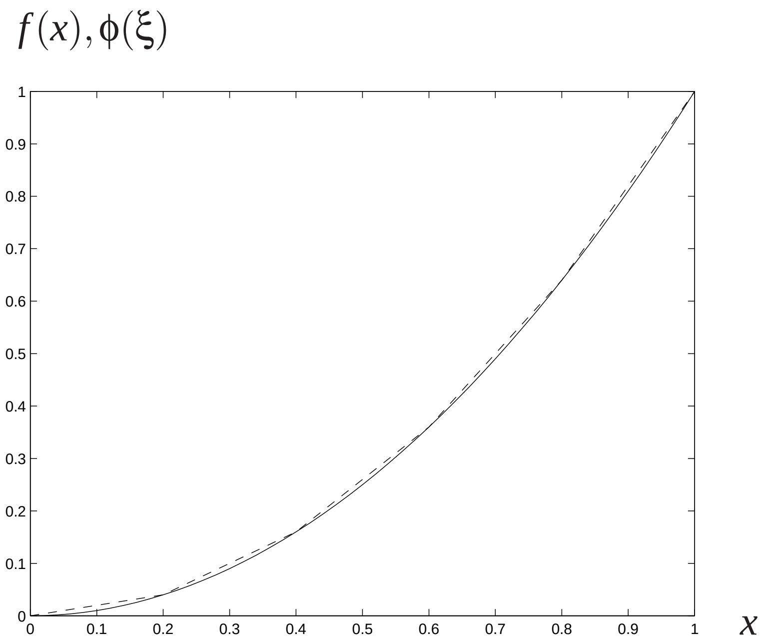


Fig. 3.13. Piece-wise linearization (shown dashed) of a function (shown solid).

Quadratic example function, continued

- For the function f illustrated in Figure 3.13:

$$\begin{aligned} d &= f(0), \\ &= 0, \\ c_j &= \frac{1}{0.2}\left(f(0.2 \times j) - f(0.2 \times (j-1))\right), \\ &= (0.4 \times j) - 0.2, j = 1, \ldots, 5. \end{aligned}$$

- To piece-wise linearize f in an optimization problem, we use ϕ as the objective instead of f, augment the decision vector to include ξ, and include the constraints that link ξ and x.
- Similarly, non-linear constraints can also be piece-wise linearized.

3.1.4.4 Smoothing

- Consider the absolute value function $|\bullet|$ defined by:

$$\forall x \in \mathbb{R}, |x| = \begin{cases} x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

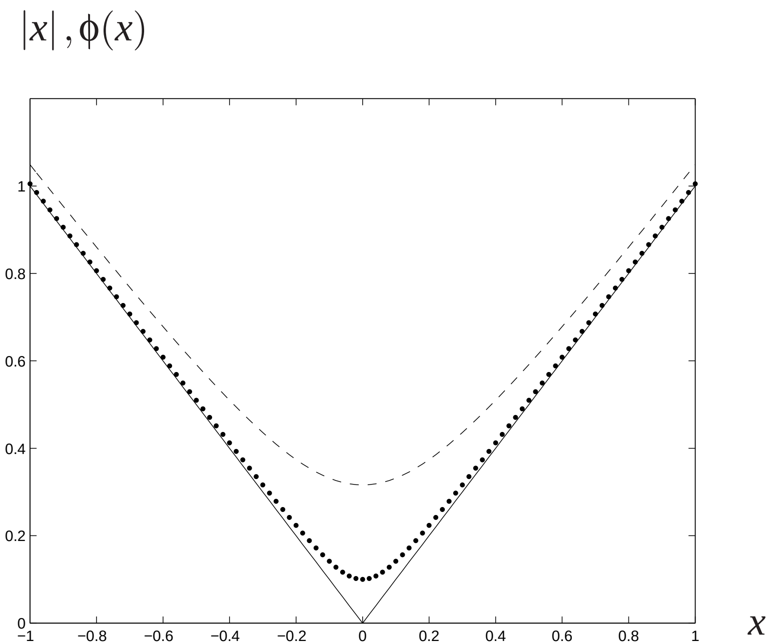


Fig. 3.14. Smoothed version for $\varepsilon = 0.1$ (shown dashed) and $\varepsilon = 0.01$ (shown dotted) of absolute value function (shown solid).

Smoothing

- This function is continuous but not differentiable.
- Consider the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}, \phi(x) = \sqrt{(|x|^2 + \varepsilon)}. \tag{3.4}$$

- We call ϕ a **smoothed version** of $|\bullet|$.
- It can be verified that for all $\varepsilon > 0$, the function ϕ is differentiable.
- Moreover, the error between ϕ and $|\bullet|$ decreases with decreasing ε.
- The smoothed function can be used as an approximation to $|\bullet|$, with a controllable approximation error determined by the choice of ε.

3.2 Variables

- The two basic techniques that we will discuss are:

  (i) **scaling**, and
  (ii) **onto transformations**.

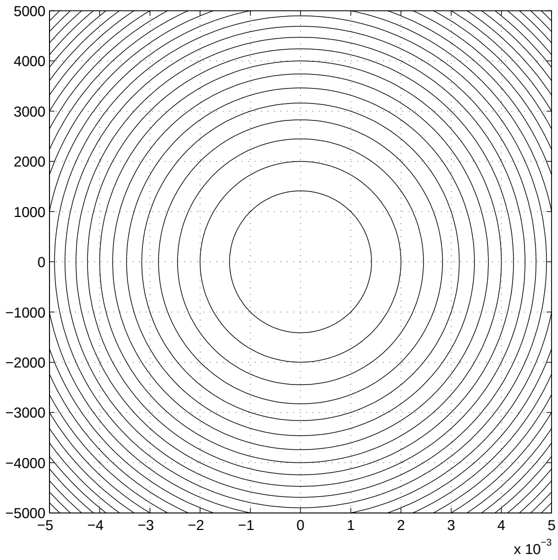
3.2.1 *Scaling*

- This simplest way to transform variables is to "scale" them.
- As a practical matter, optimization software often makes the implicit assumption that the variables have similar magnitudes at the optimum.

Example

$$\forall x \in \mathbb{R}^2, f(x) = (1000x_1)^2 + (x_2/1000)^2. \tag{3.5}$$

x_2



x_1

Fig. 3.15. Contour sets of function f defined in (3.5).

Example, continued

- If we want to obtain an solution that yields an objective that is within one unit of the minimum then we need to obtain a value of x such that (approximately):

$$\begin{aligned} |x_1^\star - x_1| &< 0.001, \\ |x_2^\star - x_2| &< 1000. \end{aligned}$$

- To appropriately weight the importance of errors in x_1 and x_2, suppose that we define scaled variables $\xi \in \mathbb{R}^2$ by:

$$\begin{aligned} \xi_1 &= 1000x_1, \\ \xi_2 &= x_2/1000. \end{aligned}$$

- Consider the objective $\phi : \mathbb{R}^2 \to \mathbb{R}$ defined by:

$$\forall \xi \in \mathbb{R}^2, \phi(\xi) = (\xi_1)^2 + (\xi_2)^2. \tag{3.6}$$

Example, continued

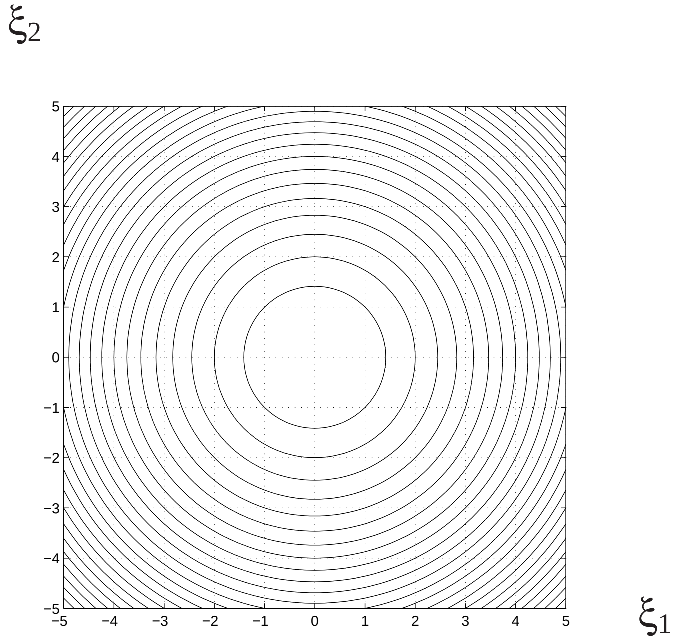


Fig. 3.16. Contour sets of function ϕ defined in (3.6) with scaled variables.

3.2.2 Onto transformations

3.2.2.1 Analysis

- We can re-write the problem in terms of new variables so long as "exploring" over the new variables also "covers" the whole of the original feasible set $\mathbb{S}$.
- This idea is embodied in the definition of an **onto** function.

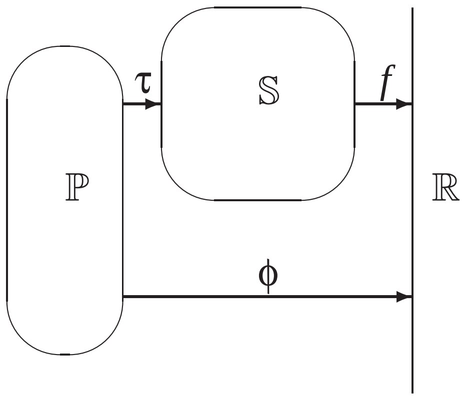


Fig. 3.17. Sets and transformations in Theorem 3.5.

Theorem 3.5 *Let* $\mathbb{S} \subseteq \mathbb{R}^n, \mathbb{P} \subseteq \mathbb{R}^{n'}, f : \mathbb{S} \to \mathbb{R}$, *let* $\tau : \mathbb{P} \to \mathbb{S}$ *be* **onto** $\mathbb{S}$, *and define* $\phi : \mathbb{P} \to \mathbb{R}$ *by:*

$$\forall \xi \in \mathbb{P}, \phi(\xi) = f(\tau(\xi)).$$

Consider the problems: $\min_{\xi \in \mathbb{P}} \phi(\xi)$ *and* $\min_{x \in \mathbb{S}} f(x)$. *Then:*

(i) *the problem* $\min_{x \in \mathbb{S}} f(x)$ *has a minimum if and only if* $\min_{\xi \in \mathbb{P}} \phi(\xi)$ *has a minimum.*

(ii) *If either one of the problems in Item* (i) *possesses a minimum (and consequently, by Item* (i)*, each one possesses a minimum), then*

$$\begin{aligned} \min_{\xi \in \mathbb{P}} \phi(\xi) &= \min_{x \in \mathbb{S}} f(x), \\ \operatorname*{argmin}_{x \in \mathbb{S}} f(x) &= \left\{ \tau(\xi) \,\middle|\, \xi \in \operatorname*{argmin}_{\xi \in \mathbb{P}} \phi(\xi) \right\}. \end{aligned}$$

□

Discussion

- To apply Theorem 3.5, it is sometimes easiest to first define a function $\tau : \mathbb{R}^{n'} \to \mathbb{R}^n$ that is onto $\mathbb{R}^n$ and then define $\mathbb{P} \subseteq \mathbb{R}^{n'}$ by:

$$\mathbb{P} = \{\xi \in \mathbb{R}^{n'} | \tau(\xi) \in \mathbb{S}\}.$$

- Then we consider the **restriction** of τ to $\mathbb{P}$.

3.2.2.2 Elimination of variables

- An important special case of Theorem 3.5 occurs when we eliminate variables.
- We first present an elementary theorem involving elimination of variables for simultaneous equations and then a corollary of Theorem 3.5 for optimization problems.

Analysis

Theorem 3.6 *Let $g : \mathbb{R}^n \to \mathbb{R}^m$, $n' \le n$ and collect the last n' entries of x together into a vector $\xi = \begin{bmatrix} x_{n-n'+1} \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n'}$. Suppose that functions $\omega_\ell : \mathbb{R}^{n'} \to \mathbb{R}$ for $\ell = 1, \ldots, (n-n')$, can be found that satisfy:*

$$\forall \begin{bmatrix} x_1 \\ \vdots \\ x_{n-n'} \\ \xi \end{bmatrix} \in \{x \in \mathbb{R}^n | g(x) = \mathbf{0}\}, \forall \ell = 1, \ldots, (n-n'), x_\ell = \omega_\ell(\xi).$$

Collect the functions $\omega_\ell, \ell = 1, \ldots, (n-n')$, into a vector function $\omega : \mathbb{R}^{n'} \to \mathbb{R}^{n-n'}$. Then, for $x \in \{x \in \mathbb{R}^n | g(x) = \mathbf{0}\}$, the vector function $\omega : \mathbb{R}^{n'} \to \mathbb{R}^{n-n'}$ expresses:

- *the sub-vector of x consisting of the first $(n-n')$ components of x,*
- *in terms of the sub-vector ξ of x consisting of the last n' components of x.*

Suppose that $\xi^\star \in \mathbb{R}^{n'}$ solves $g\left(\begin{bmatrix} \omega(\xi) \\ \xi \end{bmatrix}\right) = \mathbf{0}$. (Note that these equations involve only ξ.) Then $x^\star = \begin{bmatrix} \omega(\xi^\star) \\ \xi^\star \end{bmatrix}$ satisfies $g(x) = \mathbf{0}$.

Conversely, suppose that $x^\star \in \mathbb{R}^n$ satisfies $g(x^\star) = \mathbf{0}$. Let $\xi^\star \in \mathbb{R}^{n'}$ be the sub-vector of $x^\star$ consisting of its last n' components. Then $\xi^\star$ solves $g\left(\begin{bmatrix} \omega(\xi) \\ \xi \end{bmatrix}\right) = \mathbf{0}$. $\square$

Discussion

- In Theorem 3.6, we write the entries of g in terms of the vector ξ and the function ω by replacing $x_\ell, \ell = 1, \ldots, (n-n')$ by $\omega_\ell(\xi), \ell = 1, \ldots, (n-n')$, respectively.
- This eliminates $x_\ell, \ell = 1, \ldots, (n-n')$.
- The functions ω typically involve re-arranging some of the entries of $g(x) = \mathbf{0}$.
- In this case, we can delete the corresponding entries of g when solving $g\left(\begin{bmatrix}\omega(\xi)\\ \xi\end{bmatrix}\right) = \mathbf{0}$ since these entries are satisfied identically by $x = \begin{bmatrix}\omega(\xi)\\ \xi\end{bmatrix}$.
- The variables ξ are called the **independent variables**, while the variables $x_\ell, \ell = 1, \ldots, (n-n')$, are called the **dependent variables**.

Example

$$\forall x \in \mathbb{R}^2, g(x) = \begin{bmatrix} x_1 - x_2 \\ (x_2)^2 - x_2 \end{bmatrix}.$$

- The first entry of $g(x) = \mathbf{0}$ can be re-arranged as $x_1 = \omega_1(x_2)$, where $\xi = x_2$ and $\omega_1 : \mathbb{R} \to \mathbb{R}$ is defined by:

$$\forall x_2 \in \mathbb{R}, \omega_1(x_2) = x_2.$$

- We can delete the first entry g_1 from the equations to be solved since it is satisfied identically by $\begin{bmatrix} \omega_1(\xi) \\ \xi \end{bmatrix}$.
- We need only solve the smaller system $g_2\left(\begin{bmatrix} \omega_1(\xi) \\ \xi \end{bmatrix}\right) = 0$.

Analysis

Corollary 3.7 *Let $\mathbb{S} \subseteq \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, and $n' \leq n$ and collect the last n' entries of x together into a vector $\xi = \begin{bmatrix} x_{n-n'+1} \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n'}$. Consider the special case of the optimization problem $\min_{x \in \mathbb{S}} f(x)$ such that functions $\omega_\ell : \mathbb{R}^{n'} \to \mathbb{R}$ for $\ell = 1, \ldots, (n-n')$, can be found that satisfy:*

$$\forall \begin{bmatrix} x_1 \\ \vdots \\ x_{n-n'} \\ \xi \end{bmatrix} \in \mathbb{S}, \forall \ell = 1, \ldots, (n-n'), x_\ell = \omega_\ell(\xi).$$

(Typically, these functions correspond to $(n-n')$ of the equality constraints in the definition of $\mathbb{S}$. The condition means that these equality constraints can be re-arranged to express each of the first $n-n'$

entries of the decision vector in terms of the last n' entries.) Collect the functions $\omega_\ell, \ell = 1, \ldots, (n-n')$, into a vector function $\omega : \mathbb{R}^{n'} \to \mathbb{R}^{n-n'}$. Let $\mathbb{P} \subseteq \mathbb{R}^{n'}$ be the projection of $\mathbb{S}$ onto the last n' components of $\mathbb{R}^n$. Define $\phi : \mathbb{R}^{n'} \to \mathbb{R}$ by:

$$\forall \xi \in \mathbb{R}^{n'}, \phi(\xi) = f\left(\begin{bmatrix} \omega(\xi) \\ \xi \end{bmatrix}\right).$$

Consider the problems: $\min_{\xi\in\mathbb{P}} \phi(\xi)$ and $\min_{x\in\mathbb{S}} f(x)$. Then:

(i) *the problem $\min_{x\in\mathbb{S}} f(x)$ has a minimum if and only if $\min_{\xi\in\mathbb{P}} \phi(\xi)$ has a minimum.*

(ii) *If either one of the problems in Item (i) possesses a minimum (and consequently, by Item (i), each one possesses a minimum), then:*

$$\begin{aligned} \min_{x\in\mathbb{S}} f(x) &= \min_{\xi\in\mathbb{P}} \phi(\xi), \\ \operatorname*{argmin}_{x\in\mathbb{S}} f(x) &= \left\{ \begin{bmatrix} \omega(\xi) \\ \xi \end{bmatrix} \in \mathbb{R}^n \,\middle|\, \xi \in \operatorname*{argmin}_{\xi\in\mathbb{P}} \{\phi(\xi)\} \right\}. \end{aligned}$$

□

Example

$$\min_{x\in\mathbb{R}^2}\{(x_1-1)^2+(x_2-3)^2|x_1-x_2=0\}. \tag{3.7}$$

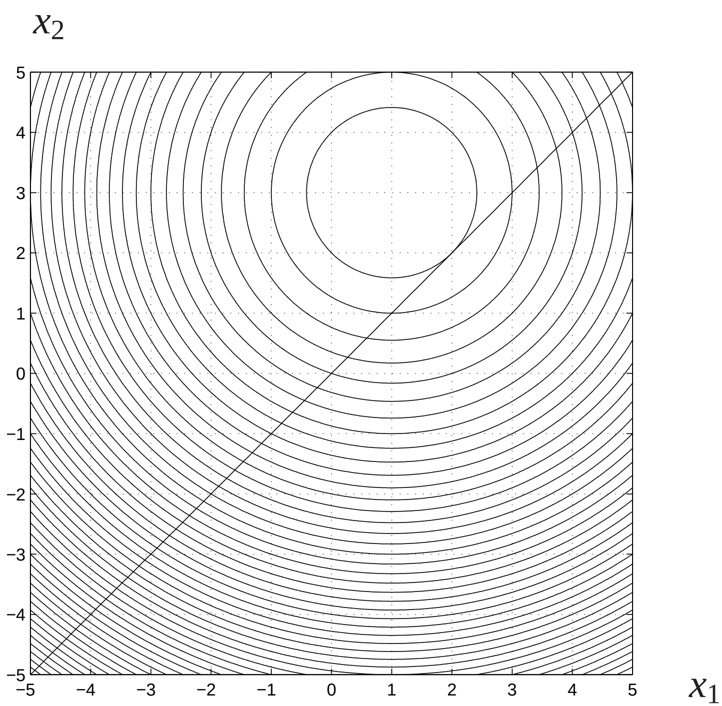


Fig. 3.18. Contour sets $\mathbb{C}_f(\tilde{f})$ of the function defined in (2.10) for values $\tilde{f} = 2,4,6,\ldots$ with feasible set superimposed. The heights of the contours decrease towards the point $\begin{bmatrix}1\\3\end{bmatrix}$.

Example, continued

- The equality constraint in this problem can be re-arranged as $x_1 = \omega_1(x_2)$, where $\xi = x_2$ and $\omega_1 : \mathbb{R} \to \mathbb{R}$ is defined by $\forall x_2 \in \mathbb{R}, \omega_1(x_2) = x_2$.
- The projection of $\mathbb{S} = \{x \in \mathbb{R}^2 | x_1 - x_2 = 0\}$ onto the last component of $\mathbb{R}^2$ is $\mathbb{P} = \mathbb{R}$.

$$\begin{aligned}
(x \in \mathbb{S}) \Rightarrow (x_1 - 1)^2 + (x_2 - 3)^2 &= (\omega_1(x_2) - 1)^2 + (x_2 - 3)^2, \\
&= (x_2 - 1)^2 + (x_2 - 3)^2, \\
&= 2(x_2)^2 - 8x_2 + 10.
\end{aligned}$$

- The transformed objective is $\phi : \mathbb{R} \to \mathbb{R}$ defined by:

$$\begin{aligned}
\forall x_2 \in \mathbb{R}, \phi(x_2) &= f\left(\begin{bmatrix} \omega(\xi) \\ \xi \end{bmatrix}\right), \\
&= f\left(\begin{bmatrix} x_2 \\ x_2 \end{bmatrix}\right), \\
&= 2(x_2)^2 - 8x_2 + 10.
\end{aligned}$$

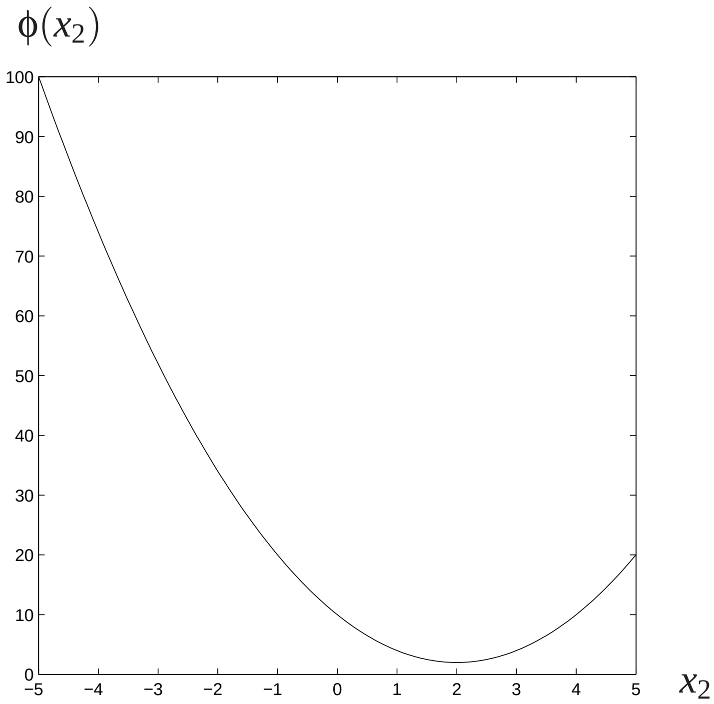


Fig. 3.19. The transformed objective function ϕ.

Example, continued

- Problem (3.7) is equivalent to:

$$\min_{x_2 \in \mathbb{R}} \{2(x_2)^2 - 8x_2 + 10\}.$$

- Inspection of Figure 3.19 yields $x_2^\star = 2$.
- The corresponding optimal value of $x_1^\star$ can be found by substituting from the eliminated constraint, according to $x_1^\star = \omega_1(x_2^\star)$.
- That is, $x_1^\star = 2$.

Discussion

- We will use elimination of variables in several places throughout the course beginning in Section 5.2.
- It is possible to generalize the idea of elimination of variables to the case where ω is not known explicitly but can only be found *implicitly*.

3.3 Constraints

- The five basic techniques we will discuss are:

  (i) **scaling** and **pre-conditioning**,
  (ii) **slack variables**,
  (iii) **changing the functional form**,
  (iv) **altering the feasible region**, and
  (v) **hierarchical decomposition**.

3.3.1 *Scaling and pre-conditioning*

- **Pre-conditioning**, involves multiplying both the coefficient matrix and the right-hand side vector on the left by a suitably chosen matrix M that:
  - does not change the set of points satisfying the constraints, but
  - makes it easier to find points satisfying the constraints.
- It is sensible to scale the entries of the constraint function so that a “significant” violation of any constraint from the perspective of the application involves roughly the same numerical value for each of the entries of the scaled constraint function.

3.3.2 Slack variables

Theorem 3.8 *Let* $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^m, h : \mathbb{R}^n \to \mathbb{R}^r$. *Consider the problems:*

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, h(x) \leq \mathbf{0} \}, \tag{3.8}$$

$$\min_{x \in \mathbb{R}^n, w \in \mathbb{R}^r} \{ f(x) | g(x) = \mathbf{0}, h(x) + w = \mathbf{0}, w \geq \mathbf{0} \}. \tag{3.9}$$

We have that:

(i) *Problem (3.8) has a minimum if and only if Problem (3.9) has a minimum.*

(ii) *If either one of the problems in Item (i) possesses a minimum (and consequently, by Item (i), each one possesses a minimum), then the minima are equal. Moreover, to each minimizer* $x^\star$ *of Problem (3.8) there corresponds a minimizer* $\begin{bmatrix} x^\star \\ w^\star \end{bmatrix}$ *of Problem (3.9) and vice versa.*

□

3.3.3 Changing the functional form

- A monotonically increasing transformation of an equality or inequality constraint function (together with the corresponding transformation of its right-hand side) does not change the feasible region, but may transform the function into being convex.

Theorem 3.9 *Let* $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $b \in \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^r$, *and* $d \in \mathbb{R}^r$. *Let* $\tau_\ell^{\nearrow} : \mathbb{R} \to \mathbb{R}, \ell = 1, \ldots, m$, *and* $\sigma_\ell^{\nearrow} : \mathbb{R} \to \mathbb{R}, \ell = 1, \ldots, r$, *each be strictly monotonically increasing and continuous on* $\mathbb{R}$. *Define* $\gamma : \mathbb{R}^n \to \mathbb{R}^m$, $\beta \in \mathbb{R}^m$, $\eta : \mathbb{R}^n \to \mathbb{R}^r$, *and* $\delta \in \mathbb{R}^r$ *by:*

$$\begin{aligned}
\forall \ell = 1, \ldots, m, \forall x \in \mathbb{R}^n, \gamma_\ell(x) &= \tau_\ell^{\nearrow}(g_\ell(x)), \\
\forall \ell = 1, \ldots, m, \beta_\ell &= \tau_\ell^{\nearrow}(b_\ell), \\
\forall \ell = 1, \ldots, r, \forall x \in \mathbb{R}^n, \eta_\ell(x) &= \sigma_\ell^{\nearrow}(h_\ell(x)), \\
\forall \ell = 1, \ldots, r, \delta_\ell &= \sigma_\ell^{\nearrow}(d_\ell).
\end{aligned}$$

Consider the problems: $\min_{x \in \mathbb{R}^n} \{f(x) | g(x) = b, h(x) \leq d\}$ *and* $\min_{x \in \mathbb{R}^n} \{f(x) | \gamma(x) = \beta, \eta(x) \leq \delta\}$. *The second problem is obtained from* *the first by transforming corresponding functions and entries of each constraint. Then:*

(i) $\min_{x\in\mathbb{R}^n}\{f(x)|g(x)=b,h(x)\leq d\}$ *has a minimum if and only if the problem* $\min_{x\in\mathbb{R}^n}\{f(x)|\gamma(x)=\beta,\eta(x)\leq\delta\}$ *has a minimum.*

(ii) *If either one of the problems in Item* (i) *possesses a minimum (and consequently, by Item* (i)*, each one possesses a minimum), then the minima are equal and they have the same minimizers.*

$\square$

- It is sometimes possible to transform a non-convex function into a convex function.
- This applies in the case of a **posynomial function**:

Definition 3.1 Let $A\in\mathbb{R}^{m\times n}$ and $B\in\mathbb{R}^m_{++}$ and define $f:\mathbb{R}^n_{++}\to\mathbb{R}$ by:

$$\forall x\in\mathbb{R}^n_{++}, f(x)=\sum_{\ell=1}^{m}B_\ell(x_1)^{A_{\ell 1}}(x_2)^{A_{\ell 2}}\cdots(x_n)^{A_{\ell n}}.$$

The function f is called a **posynomial function**. If $m=1$ then f is called a **monomial function**. $\square$

3.3.4 Altering the feasible region

Theorem 3.10 *Let* $\underline{\mathbb{S}} \subseteq \mathbb{S} \subseteq \overline{\mathbb{S}} \subseteq \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ *and consider the problems:*

$$\min_{x \in \underline{\mathbb{S}}} f(x), \min_{x \in \mathbb{S}} f(x), \min_{x \in \overline{\mathbb{S}}} f(x),$$

and suppose that they all have minima and minimizers. Then:

(i)

$$\min_{x \in \underline{\mathbb{S}}} f(x) \geq \min_{x \in \mathbb{S}} f(x) \geq \min_{x \in \overline{\mathbb{S}}} f(x).$$

(ii) *If* $x^\star \in \operatorname{argmin}_{x \in \overline{\mathbb{S}}} f(x)$ *and* $x^\star \in \mathbb{S}$ *then* $\min_{x \in \mathbb{S}} f(x) = \min_{x \in \overline{\mathbb{S}}} f(x)$ *and* $\operatorname{argmin}_{x \in \mathbb{S}} f(x) = \left(\operatorname{argmin}_{x \in \overline{\mathbb{S}}} f(x)\right) \cap \mathbb{S}$.

(iii) *If* $x^\star \in \operatorname{argmin}_{x \in \mathbb{S}} f(x)$ *and* $x^\star \in \underline{\mathbb{S}}$ *then* $\min_{x \in \mathbb{S}} f(x) = \min_{x \in \underline{\mathbb{S}}} f(x)$ *and* $\operatorname{argmin}_{x \in \underline{\mathbb{S}}} f(x) = (\operatorname{argmin}_{x \in \mathbb{S}} f(x)) \cap \underline{\mathbb{S}}$.

□

3.3.4.1 *Enlarging or relaxing the feasible set*

- The problem $\min_{x \in \overline{\mathbb{S}}} f(x)$ is called a **relaxation** of or a **relaxed** version of the original problem $\min_{x \in \mathbb{S}} f(x)$.
- If the minimizer of the relaxed problem $\min_{x \in \overline{\mathbb{S}}} f(x)$ happens to lie in $\mathbb{S}$, then a minimizer of the original problem $\min_{x \in \mathbb{S}} f(x)$ has been found.
- It is sometimes easier to optimize over a larger set than a smaller set, if the larger set has a more suitable structure.

  (i) $\overline{\mathbb{S}}$ is convex while $\mathbb{S}$ is not, and

  (ii) $\overline{\mathbb{S}}$ involves temporarily ignoring some of the constraints, yielding an easier problem.

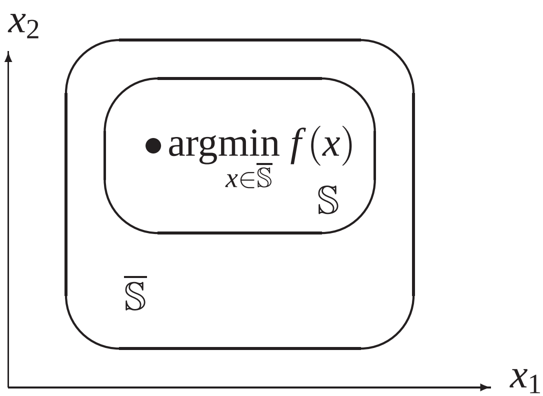


Fig. 3.20. Illustration of relaxing the feasible set.

3.3.4.2 *Constricting the feasible set*

- Item (iii) in Theorem 3.10 simply formalizes a way to use *a priori* knowledge to narrow a search: if an optimizer is known to lie in a subset $\underline{\mathbb{S}}$ of $\mathbb{S}$ then we can confine our search to that subset.
- This can be useful if it is easier to search over $\underline{\mathbb{S}}$ than over $\mathbb{S}$.

3.3.4.3 Divide and conquer

- We can generalize the idea of constricting the feasible set to develop a **divide and conquer** approach.
- Suppose that $\mathbb{S}_1 \subseteq \mathbb{S}$, $\mathbb{S}_2 \subseteq \mathbb{S}$, and $\mathbb{S}_1 \cup \mathbb{S}_2 = \mathbb{S}$.
- If the minimizer of the problem over $\mathbb{S}$ exists, then it must be contained in either $\mathbb{S}_1$ or $\mathbb{S}_2$, (or both).
- We solve both $\min_{x \in \mathbb{S}_1} f(x)$ and $\min_{x \in \mathbb{S}_2} f(x)$ and check for the smaller minimum and corresponding minimizer.
- This yields the minimum and minimizer of $\min_{x \in \mathbb{S}} f(x)$.

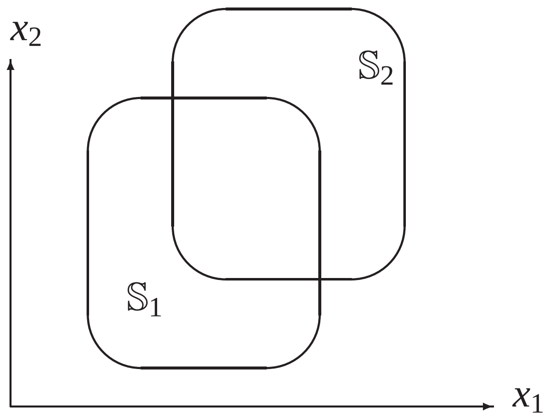


Fig. 3.21. Illustration of divide and conquer.

3.3.5 Hierarchical decomposition

- Consider a feasible set $\mathbb{S} \subseteq \mathbb{R}^{n+s}$ such that:

$$\mathbb{S} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^{n+s} \middle| x \in \mathbb{S}_1, y \in \mathbb{S}_2(x) \right\},$$

- where $\mathbb{S}_1 \subseteq \mathbb{R}^n$ and $\mathbb{S}_2 : \mathbb{S}_1 \to (2)^{(\mathbb{R}^s)}$ is a set-valued function.

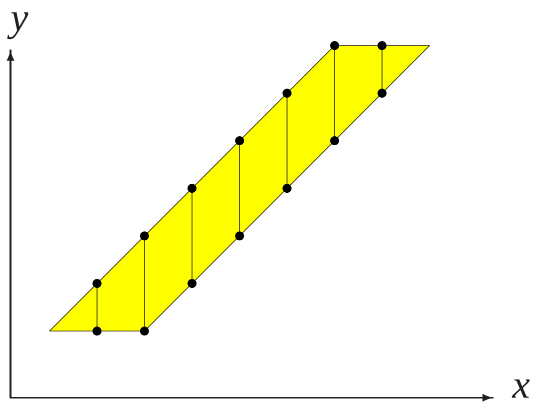


Fig. 3.22. Illustration of hierarchical decomposition.

Theorem 3.11 *Suppose that* $\mathbb{S} \subseteq \mathbb{R}^{n+s}$ *is of the form:*

$$\mathbb{S} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^{n+s} \middle| x \in \mathbb{S}_1, y \in \mathbb{S}_2(x) \right\},$$

with $\mathbb{S}_1 \subseteq \mathbb{R}^n$ *and, for each* $x \in \mathbb{S}_1$, $\mathbb{S}_2(x) \subseteq \mathbb{R}^s$*. Let* $f : \mathbb{S} \to \mathbb{R}$ *and suppose that, for each* $x \in \mathbb{S}_1$*, the minimization problem* $\min_{y \in \mathbb{S}_2(x)} f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$ *has a minimum. Consider the problems:*

$$\min_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \in \mathbb{S}} f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \text{ and } \min_{x \in \mathbb{S}_1} \left\{ \min_{y \in \mathbb{S}_2(x)} f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \right\}.$$

Then:

(i) $\min_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \in \mathbb{S}} f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$ *has a minimum if and only if*

$\min_{x \in \mathbb{S}_1} \left\{ \min_{y \in \mathbb{S}_2(x)} f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \right\}$ *has a minimum.*

(ii) *If either one of the problems in Item* (i) *possesses a minimum (and consequently, by Item* (i)*, each one possesses a minimum), then:*

$$\min_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]\in\mathbb{S}} f\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \min_{x\in\mathbb{S}_1}\left\{\min_{y\in\mathbb{S}_2(x)} f\left(\begin{bmatrix}x\\y\end{bmatrix}\right)\right\},$$

$$\operatorname*{argmin}_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]\in\mathbb{S}} f\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \left\{\begin{bmatrix}\hat{x}\\\hat{y}\end{bmatrix}\in\mathbb{R}^{n+s}\;\middle|\;\begin{array}{c}\hat{x}\in\operatorname*{argmin}_{x\in\mathbb{S}_1}\left\{\min_{y\in\mathbb{S}_2(x)} f\left(\begin{bmatrix}x\\y\end{bmatrix}\right)\right\},\\ \hat{y}\in\arg\min_{y\in\mathbb{S}_2(\hat{x})} f\left(\begin{bmatrix}\hat{x}\\y\end{bmatrix}\right)\end{array}\right\}.$$

$\square$

- Theorem 3.11 allows us to hold some of the decision vector *constant* temporarily while we optimize over the rest of the decision vector.
- We keep $x \in \mathbb{S}_1$ constant *temporarily* or think of it as a parameter while we optimize the **inner problem** over $y \in \mathbb{S}_2(x)$.
- If we can solve for the solution of the inner problem as a function of x, or can approximate its dependence on x, then we can use this functional dependence in the **outer problem**.

Example

- Consider the feasible set:

$$\mathbb{S} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \middle| (x)^2 + (y)^2 = 1 \right\},$$

- which is the set of points on the unit circle in the plane.
- We can re-write this set in the form:

$$\mathbb{S} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| -1 \leq x \leq 1, y \in \left\{ \sqrt{1-(x)^2}, -\sqrt{1-(x)^2} \right\} \right\},$$

- where $\mathbb{S}_1 = \{x \in \mathbb{R} | -1 \leq x \leq 1\}$ is the projection of $\mathbb{S}$ onto the first component of $\mathbb{R}^2$.
- In this case, for each $x \in \mathbb{S}_1$, the inner minimization problem in Theorem 3.11 involves finding the minimum over a set with just two elements, namely $\mathbb{S}_2(x) = \{\sqrt{1-(x)^2}, -\sqrt{1-(x)^2}\}$.
- Even if the objective is non-convex, and despite the fact that $\mathbb{S}_2(x)$ is not a convex set, it may be easy to perform this minimization.

Discussion

- If $\mathbb{S}$ is convex and f is a convex function on $\mathbb{S}$ then both the inner problem and the outer problem are convex.
- Hierarchical decomposition is also useful when holding $x \in \mathbb{S}_1$ constant yields an inner problem with a particular structure that is easy to solve or for which a convenient approximate solution is possible.
  - This leads to **Bender's decomposition**.

3.4 Duality

- Taking the **dual** of a problem is a process whereby a new problem is defined where the role of the variables and the constraints is either partially or completely exchanged.
- Let $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^r$.
- Consider the problem:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, h(x) \leq \mathbf{0} \}. \tag{3.10}$$

- We define two functions associated with f, g, and h, called the **Lagrangian** and the **dual function.**
- We then consider the relationship between these functions and minimizing f.

3.4.1 Lagrangian

Definition 3.2 Consider the function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^\dagger g(x) + \mu^\dagger h(x). \quad (3.11)$$

The function $\mathcal{L}$ is called the **Lagrangian** and the variables λ and μ are called the **dual variables**. If there are no equality constraints then $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ is defined by omitting the term $\lambda^\dagger g(x)$ from the definition, while if there are no inequality constraints then $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is defined by omitting the term $\mu^\dagger h(x)$ from the definition. $\Box$

- Sometimes, the symbol for the dual variables is introduced when the problem is defined by writing it in parenthesis after the constraint, as in the following:

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } g(x) = \mathbf{0}, \qquad (\lambda).$$

3.4.2 Dual function

- Associated with the Lagrangian, we make:

Definition 3.3 Consider the function $\mathcal{D} : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R} \cup \{-\infty\}$ defined by:

$$\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r}, \mathcal{D}(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu). \tag{3.12}$$

The function $\mathcal{D}$ is called the **dual function.** It is an extended real function. If there are no equality constraints or there are no inequality constraints, respectively, then the dual function $\mathcal{D} : \mathbb{R}^r \to \mathbb{R} \cup \{-\infty\}$ or $\mathcal{D} : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$ is defined in terms of the corresponding Lagrangian. The set of points on which the dual function takes on real values is called the **effective domain** $\mathbb{E}$ of the dual function:

$$\mathbb{E} = \left\{ \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r} \middle| \mathcal{D}(\lambda, \mu) > -\infty \right\}.$$

The restriction of $\mathcal{D}$ to $\mathbb{E}$ is a real-valued function $\mathcal{D} : \mathbb{E} \to \mathbb{R}$. $\square$

Discussion

- Recall Definition 2.17 of a concave function.
- The usefulness of the dual function stems in part from the following:

Theorem 3.12 *Let $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^r$. Consider the corresponding Lagrangian defined in (3.11), the dual function defined in (3.12), and the effective domain $\mathbb{E}$ of the dual function. The effective domain $\mathbb{E}$ of the dual function is a convex set. The dual function is concave on $\mathbb{E}$.* $\square$

- The convexity of the effective domain and the concavity of the dual function on the effective domain does not depend on any property of the objective nor of the constraint functions.

3.4.3 Dual problem

Theorem 3.13 *Let* $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^m$*, and* $h : \mathbb{R}^n \to \mathbb{R}^r$*. Let* $\lambda \in \mathbb{R}^m$ *and* $\mu \in \mathbb{R}^r_+$ *and suppose that* $\hat{x} \in \{x \in \mathbb{R}^n | g(x) = \mathbf{0}, h(x) \leq \mathbf{0}\}$*. That is,* $\hat{x}$ *is feasible for Problem (3.10). Then:*

$$f(\hat{x}) \geq \mathcal{D}(\lambda, \mu), \tag{3.13}$$

where $\mathcal{D} : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R} \cup \{-\infty\}$ *is the dual function defined in (3.12).*

Proof By definition of $\mathcal{D}$,

$$\begin{aligned}
\mathcal{D}(\lambda, \mu) &= \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu), \\
&= \inf_{x \in \mathbb{R}^n} \{f(x) + \lambda^\dagger g(x) + \mu^\dagger h(x)\}, \text{ by definition of } \mathcal{L}, \\
&\leq f(\hat{x}) + \lambda^\dagger g(\hat{x}) + \mu^\dagger h(\hat{x}), \text{by definition of inf}, \\
&\leq f(\hat{x}),
\end{aligned}$$

since $g(\hat{x}) = \mathbf{0}$, $h(\hat{x}) \leq \mathbf{0}$, and $\mu \geq \mathbf{0}$. $\square$

Discussion

- Theorem 3.13 enables us to gauge whether we are close to a minimum of Problem (3.10).
- For any value of $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^r_+$, we know that the minimum of Problem (3.10) is no smaller than $\mathcal{D}(\lambda, \mu)$.
- This lower bound will be incorporated into a stopping criterion for iterative algorithms.

Corollary 3.14 *Let $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^r$. Then:*

$$\begin{aligned}
\inf_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, h(x) \leq \mathbf{0} \} &\geq \sup_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{R}^{m+r}} \{ \mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0} \}, \\
&= \sup_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{E}} \{ \mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0} \},
\end{aligned}$$

where $\mathbb{E}$ is the effective domain of $\mathcal{D}$. Moreover, if Problem (3.10) has a minimum then:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, h(x) \leq \mathbf{0} \} \geq \sup_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{E}} \{ \mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0} \}. \qquad (3.14)$$

If Problem (3.10) is unbounded below then:

$$\forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r_+, \mathcal{D}(\lambda, \mu) = -\infty,$$

so that $\mathbb{E}_+ = \left\{ \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{E} \middle| \mu \geq \mathbf{0} \right\} = \emptyset.$

If the problem $\sup_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{R}^{m+r}} \{ \mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0} \}$ *is unbounded above then Problem (3.10) is infeasible.* $\square$

Discussion

- This result is called **weak duality**.
- The right-hand side of (3.14) is called the **dual problem.**
- If $\mathbb{E}_+ = \emptyset$ we say that the dual problem is infeasible.
- The inequality in (3.14) can be strict, in which case the difference between the left and right-hand sides is called the **duality gap**.
- If the left and right sides are the same, we say that there is no duality gap or that the duality gap is zero.
- Evaluating the right-hand side of (3.14) requires:
  - evaluating the dependence of the infimum of the **inner problem** $\inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu)$ in the definition of $\mathcal{D}$ as a function of λ and μ,
  - finding the supremum of the **outer problem** $\sup_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right] \in \mathbb{E}} \{\mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0}\}$.

Discussion, continued

- In some circumstances, the inequality in (3.14) can be replaced by equality and the sup and inf can be replaced by max and min, so that the right-hand side of (3.14) equals the minimum of Problem (3.10) and the right-hand side becomes:

$$\max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}}\{\mathcal{D}(\lambda,\mu)|\mu\geq\mathbf{0}\} = \max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}}\left\{\min_{x\in\mathbb{R}^n}\{f(x)+\lambda^\dagger g(x)+\mu^\dagger h(x)\}\,\middle|\,\mu\geq\mathbf{0}\right\}, \tag{3.15}$$

- having an inner minimization problem embedded in an outer maximization problem.
- By Theorem 3.12, $\mathcal{D}$ is concave on $\mathbb{E}$, so that, by Theorem 2.4, it has at most one local maximum.

Discussion, continued

- The dual formulation provides a useful transformation if:
  - the dual problem has maximum equal to the minimum of the primal problem, and
  - the minimizer of the inner problem in the definition of the dual function sheds light on the minimizer of the primal problem,
- The requirements for these conditions to hold depend on the convexity of the primal problem and on other technical conditions on the functions, which we will discuss in detail in Parts IV and V.
- In the next section, we will consider an example where such conditions happen to hold.

3.4.4 Example

- Consider the problem $\min_{x\in\mathbb{R}}\{f(x)|g(x)=0\}$ where $f:\mathbb{R}\to\mathbb{R}$ and where $g:\mathbb{R}\to\mathbb{R}$ are defined by:

$$\begin{aligned}\forall x\in\mathbb{R}, f(x) &= (x)^2,\\ \forall x\in\mathbb{R}, g(x) &= 3-x.\end{aligned}$$

- Since there are no inequality constraints, we will omit the argument μ of $\mathcal{L}$ and of $\mathcal{D}$.
- We consider the dual function $\mathcal{D}:\mathbb{R}\to\mathbb{R}\cup\{-\infty\}$ defined by:

$$\begin{aligned}\forall\lambda\in\mathbb{R}, \mathcal{D}(\lambda) &= \inf_{x\in\mathbb{R}} \mathcal{L}(x,\lambda),\\ &= \inf_{x\in\mathbb{R}}\{(x)^2+\lambda(3-x)\},\\ &= \inf_{x\in\mathbb{R}}\left\{\left(x-\frac{\lambda}{2}\right)^2+3\lambda-\frac{(\lambda)^2}{4}\right\},\\ &= 3\lambda-\frac{(\lambda)^2}{4}.\end{aligned}$$

Example, continued

- Therefore, $\mathbb{E} = \mathbb{R}$ and since $\mathcal{D}$ is quadratic and strictly concave, the dual problem has a maximum and:

$$\begin{aligned}
\max_{\lambda \in \mathbb{E}} \{\mathcal{D}(\lambda)\} &= \max_{\lambda \in \mathbb{R}} \left\{ 3\lambda - \frac{(\lambda)^2}{4} \right\}, \\
&= \max_{\lambda \in \mathbb{R}} \left\{ -\left(\frac{\lambda}{2} - 3\right)^2 + 9 \right\}, \\
&= 9,
\end{aligned}$$

- with maximizer $\lambda^\star = 6$.
- The value of the minimizer of $\mathcal{L}(\bullet, \lambda^\star)$ is $x^\star = \frac{\lambda^\star}{2} = 3$, which is the minimizer of the equality-constrained problem.
- We have solved the primal equality-constrained problem by solving the dual problem.
- There is no duality gap.

3.4.5 Discussion

- To understand the Lagrangian, consider $f_\mathrm{p} : \mathbb{R}^n \to \mathbb{R}_+ \cup \{\infty\}$ defined by:

$$\forall x \in \mathbb{R}^n, f_\mathrm{p}(x) = \sup_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r_+} \{\lambda^\dagger g(x) + \mu^\dagger h(x)\}.$$

- f_p is a discontinuous penalty function for the constraints $g(x) = \mathbf{0}$ and $h(x) \leq \mathbf{0}$, since:
  - if $g(x) = \mathbf{0}$ and $h(x) \leq \mathbf{0}$, then $\mu \geq \mathbf{0}$ implies $\lambda^\dagger g(x) + \mu^\dagger h(x) \leq 0$, but $\mathbf{0}^\dagger g(x) + \mathbf{0}^\dagger h(x) = 0$, so $f_\mathrm{p}(x) = 0$, whereas
  - if $g_\ell(x) \neq 0$ or $h_\ell(x) > 0$ then we can make $\lambda^\dagger g(x) + \mu^\dagger h(x)$ arbitrarily large by choosing λ_ℓ and μ_ℓ appropriately, so $f_\mathrm{p}(x) = \infty$.

Discussion, continued

- Now note that:

$$
\begin{aligned}
\forall x \in \mathbb{R}^n, f(x) + f_{\mathrm{p}}(x) &= f(x) + \sup_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r_+} \{\lambda^\dagger g(x) + \mu^\dagger h(x)\}, \\
&= \sup_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r_+} \{f(x) + \lambda^\dagger g(x) + \mu^\dagger h(x)\}, \\
&= \sup_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r_+} \{\mathcal{L}(x, \lambda, \mu)\},
\end{aligned}
$$

- so that the terms in the Lagrangian provide a penalty function for the constraints when λ and $\mu \geq \mathbf{0}$ are chosen appropriately.

Discussion, continued

- For each equality constraint $g_\ell(x) = 0$ in the primal problem we have created a new variable λ_ℓ in the dual problem.
- For each inequality constraint $h_\ell(x) \leq 0$ in the primal problem we have created a new variable μ_ℓ and a new constraint $\mu_\ell \geq 0$ in the dual problem.
- In some circumstances, such as the example in Section 3.4.4:
  - the minimization over $x \in \mathbb{R}^n$ in the inner problem in (3.15) can be performed analytically or particularly easily numerically, or
  - each entry x_k can be eliminated,
- making the inner problem easy to solve.

3.5 Summary

- These transformations involved:

  (i) the objective,
  (ii) the variables,
  (iii) the constraints, and
  (iv) duality.