Course notes for EE394V Restructured Electricity Markets: Locational Marginal Pricing

Ross Baldick

www.ece.utexas.edu/~baldick/classes/394V/EE394V.html

4
Optimization

(i) Basic definitions,
(ii) Duality,
(iii) Continuous unconstrained problems,
(iv) Continuous equality-constrained problems,
(v) Continuous linear inequality-constrained problems,
(vi) Continuous non-linear inequality-constrained problems,
(vii) Integer problems,

(viii) Mixed-integer problems,
(ix) Homework exercises.

4.1 Basic definitions

4.1.1 Decision vector

- As with solution of simultaneous equations, the optimization problems we consider will involve choices of a value of a **decision vector**.
- We will usually denote the decision vector by x, y, or z.
- It will be chosen from $\mathbb{R}^n$ or from some subset $\mathbb{S}$ of $\mathbb{R}^n$.
- As previously, entries of the decision vector will be indexed by subscripts.
- Example:
  - the choice of dispatch for generator k is x_k, while
  - the choices of dispatch for all generators is x.

4.1.2 Objective

- Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ 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**.
- Examples:
  - the operating cost of a generator, and
  - the sum of the operating costs of all generators in a system.

4.1.2.1 Example

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

$$\forall x \in \mathbb{R}^2, f(x) = (x_1)^2 + (x_2)^2 + 2x_2 - 3. \tag{4.1}$$

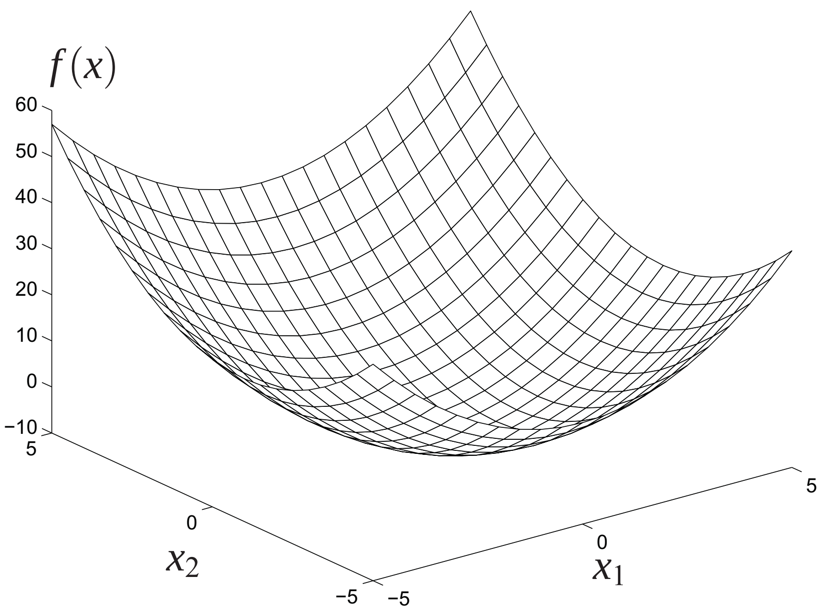

Fig. 4.1. Graph of the example objective function defined in (4.1).

4.1.2.2 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 4.1.14 once we have discussed optimization in more detail.

4.1.3 Feasible set

- Our problems will typically 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.
- Examples:
  - the set representing the allowable range of operating conditions of a generator, and
  - the set representing the allowable range of operating conditions for all the generators in a system.

4.1.4 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 4.1 Let $\mathbb{S} \subseteq \mathbb{R}^n$, $f : \mathbb{S} \to \mathbb{R}$, and $f^\star \in \mathbb{R}$. Then by:

$$f^\star = \min_{x \in \mathbb{S}} f(x), \tag{4.2}$$

we mean that:

$$\exists x^\star \in \mathbb{S} \text{ such that: } (f^\star = f(x^\star)) \text{ and } ((x \in \mathbb{S}) \Rightarrow (f(x^\star) \leq f(x))). \tag{4.3}$$

$\square$

- We say that $f^\star$ is the minimum of $f(x)$ over values of x in the set $\mathbb{S}$ or that $f^\star$ is the minimum of $f(x)$ such that $x \in \mathbb{S}$.
- Example:
  - find the choices of dispatch x for all generators that minimizes the sum of the operating costs and such that the dispatch meets demand and is within the allowable operating conditions for all generators.

4.1.5 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:

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

- where $\emptyset$ is the empty set.
- 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 $\min_{x\in\mathbb{S}} f(x)$.

4.1.6 Level and contour sets

Definition 4.2 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:

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

□

- Contour and level sets are useful for visualizing functions.
- If $f : \mathbb{R}^3 \to \mathbb{R}$, how many dimensions are needed to graph f? How many are needed to show the contour set?

4.1.6.1 Example

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

$$\forall x \in \mathbb{R}^2, f(x) = (x_1 - 1)^2 + (x_2 - 3)^2. \tag{4.4}$$

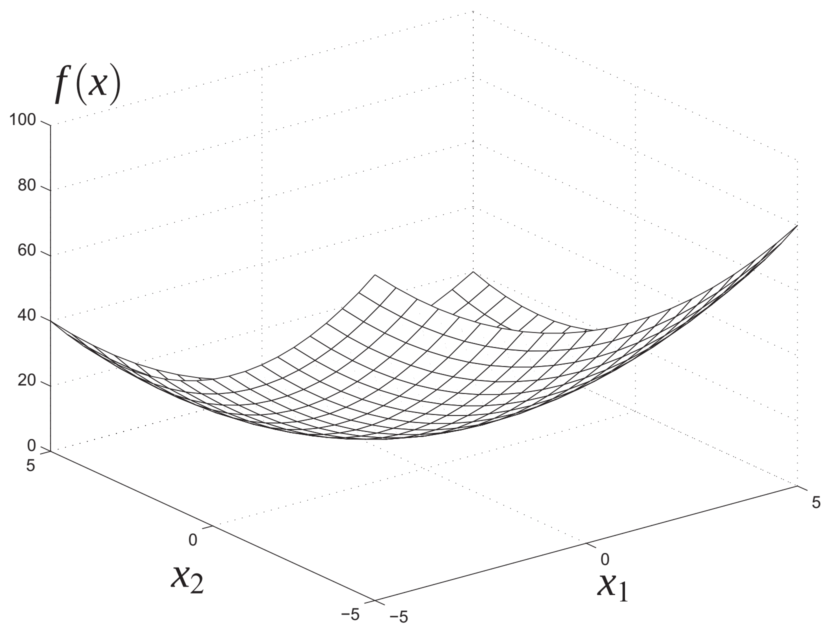

Fig. 4.2. Graph of function defined in (4.4).

4.1.6.2 Contour set for example

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

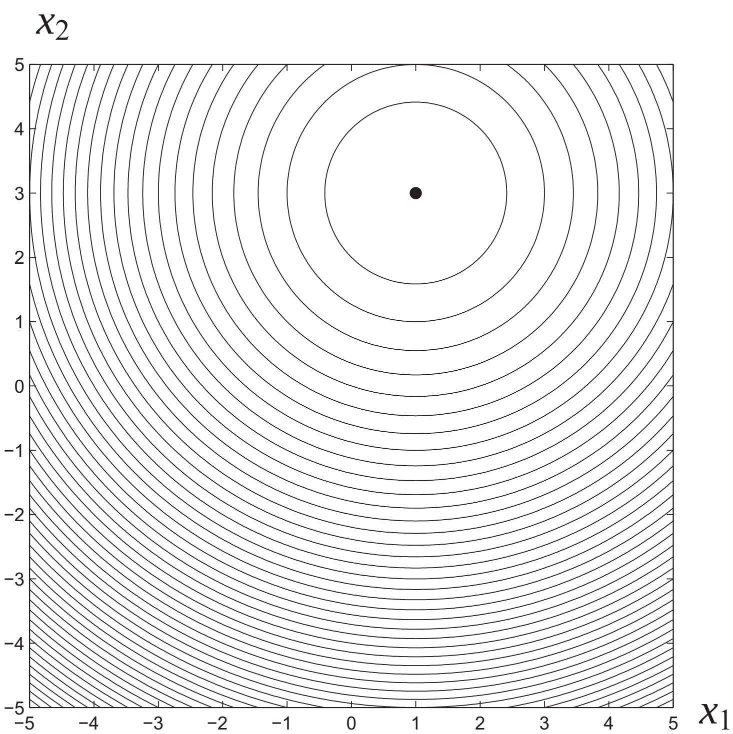

Fig. 4.3. Contour sets $\mathbb{C}_f(\tilde{f})$ of the function defined in (4.4) for values $\tilde{f} = 0, 2, 4, 6, \ldots$. The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$, which is illustrated with a $\bullet$ and is the contour of height 0.

4.1.7 Unconstrained optimization

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

4.1.7.1 Example

- For example, consider the objective $f : \mathbb{R}^2 \to \mathbb{R}$ defined in (4.4):

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

- From Figure 4.3, 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.

4.1.8 Equality-constrained optimization

- If $g : \mathbb{R}^n \to \mathbb{R}^m$ and the feasible set is $\mathbb{S} = \{x \in \mathbb{R}^n | g(x) = \mathbf{0}\}$ then the problem is said to be **equality-constrained.**
- The power flow equations in Section 3.2.7 is an example of equality constraints.

4.1.8.1 Sub-types of equality-constrained optimization problems

Linearly constrained

- If g is affine then the problem is called **linearly constrained**.
- The DC power flow approximation to the power flow equations in Section 3.6.2 is an example of linear equality constraints.

Example

$$\begin{aligned} \forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\ \forall x \in \mathbb{R}^2, g(x) &= x_1 - x_2, \\ \min_{x \in \mathbb{R}^2} \{f(x) | g(x) = 0\} &= \min_{x \in \mathbb{R}^2} \{f(x) | x_1 - x_2 = 0\}. \end{aligned} \quad (4.5)$$

Example, continued

- The unique minimizer of Problem (4.5) is $x^\star = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

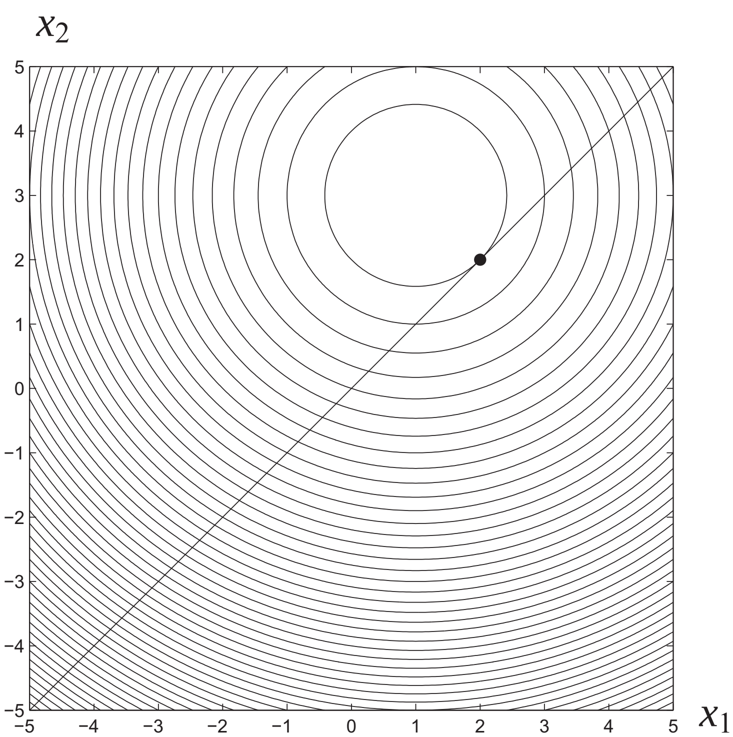

Fig. 4.4. Contour sets $\mathbb{C}_f(\tilde{f})$ of function repeated from Figure 4.3 with feasible set from Problem (4.5) superimposed. The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. The minimizer $x^\star = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ is illustrated with a $\bullet$.

Non-linearly constrained

- If there is no restriction on g then the problem is called **non-linearly constrained**.
- The AC power flow equations in Section 3.2.7 is an example of non-linear equality constraints.

Example

- For example, consider the same objective as previously, $f : \mathbb{R}^2 \to \mathbb{R}$ defined in (4.4):
$$\forall x \in \mathbb{R}^2, f(x) = (x_1 - 1)^2 + (x_2 - 3)^2.$$
- 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:
$$\min_{x \in \mathbb{R}^2} \{ f(x) | g(x) = 0 \}. \tag{4.6}$$

Example, continued

- The unique minimizer of Problem (4.6) is $x^\star \approx \begin{bmatrix} 0.5 \\ 0.9 \end{bmatrix}$.

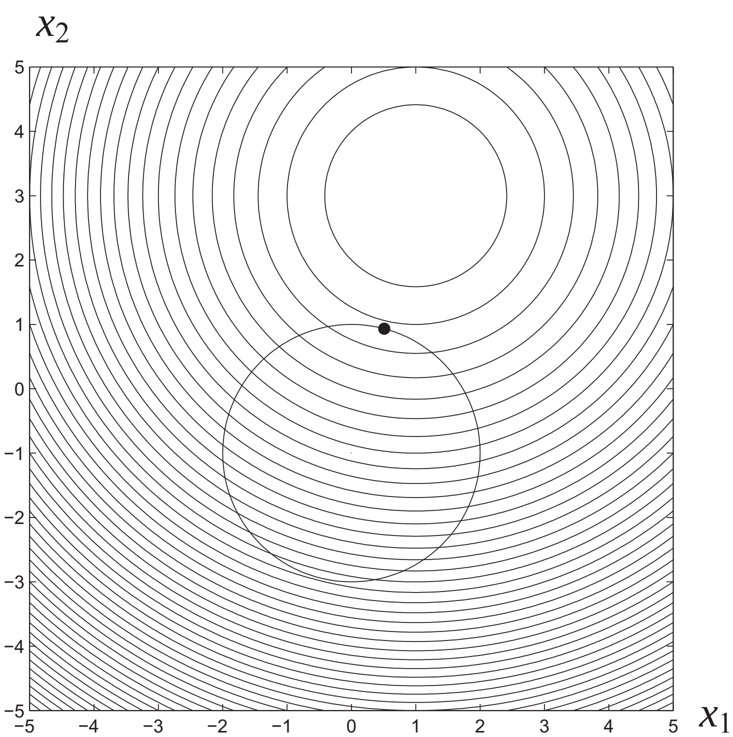

Fig. 4.5. Contour sets $\mathbb{C}_f(\tilde{f})$ of function repeated from Figure 4.3 with feasible set from Problem (4.6) superimposed. The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. The minimizer $x^\star$ is illustrated as a ●.

4.1.9 *Inequality-constrained optimization*

- If $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^r$, and the feasible set is $\mathbb{S} = \{x \in \mathbb{R}^n | g(x) = \mathbf{0}, h(x) \leq \mathbf{0}\}$ then the problem is said to be **inequality-constrained.**
- The line flow constraints in Section 3.7.2 is an example of inequality constraints.

4.1.9.1 *Sub-types of inequality-constrained optimization problems*

Linear inequality constraints

- If h is affine then the problem is **linear inequality-constrained**.
- The DC power flow approximation to line flow constraints in Section 3.7.3 is an example of linear inequality constraints.

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

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

$$\min_{x \in \mathbb{R}^2} \{ f(x) | g(x) = 0, h(x) \leq \mathbf{0} \} = \min_{x \in \mathbb{R}^2} \{ x_1 - x_2 | x_1 + x_2 - 1 = 0, x_1 \geq 0, x_2 \geq 0 \}. \tag{4.7}$$

Example, continued

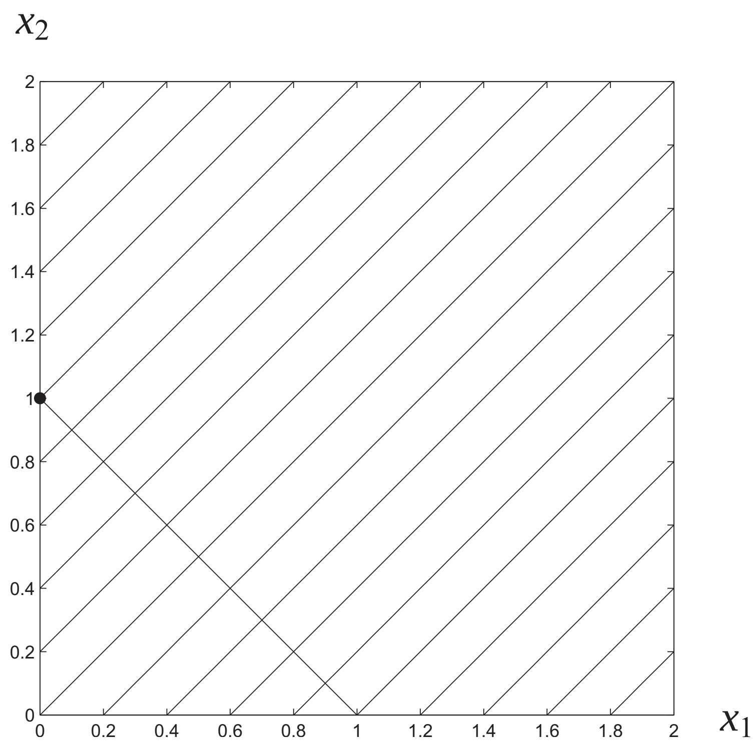

Fig. 4.6. Contour sets $\mathbb{C}_f(\tilde{f})$ of objective function and feasible set for Problem (4.7). The contour sets are the parallel lines. The feasible set is shown as the line joining the two points $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. The heights of the contours decrease to the left and up. The minimizer $x^\star = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is illustrated as a $\bullet$.

Linear program, continued

- 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 $d \in \mathbb{R}^r$ and where $c^\dagger$ is the **transpose** of c.
- For Problem (4.7), the appropriate vectors and matrices are:

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

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

$$\min_{x\in\mathbb{R}^2}\{c^\dagger x | Ax = b, x \geq \mathbf{0}\}. \tag{4.8}$$

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**.
- The reliability and capabilities of commercial linear programming packages are two of several reasons why linearized approximations to the the power flow equality and inequality constraints are typically used in practice in implementation of electricity markets as optimization problems.

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**.
- Some electricity market formulations are most naturally expressed as a quadratic program:
  - may actually be solved through a linearized approximation.

Example

$$\begin{aligned}\forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\ \forall x \in \mathbb{R}^2, g(x) &= x_1 - x_2, \\ \forall x \in \mathbb{R}^2, h(x) &= 3 - x_2.\end{aligned} \tag{4.9}$$

Example, continued

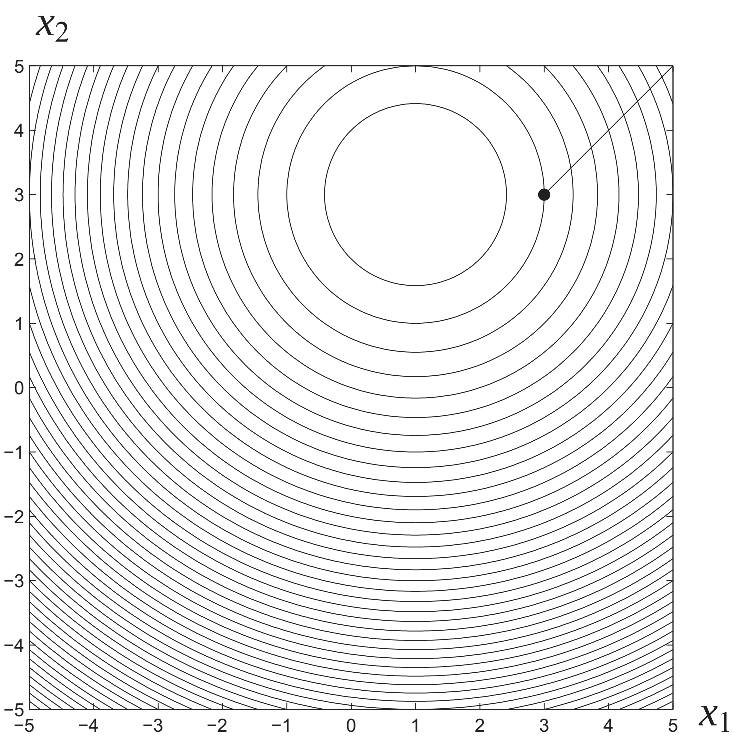

Fig. 4.7. Contour sets $\mathbb{C}_f(\tilde{f})$ of objective function and feasible set for Problem (4.10). The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. The feasible set is the "half-line" starting at the point $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$. The minimizer $x^\star = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ is illustrated with a $\bullet$.

Example, continued

$$\min_{x \in \mathbb{R}^2} \{f(x) | g(x) = 0, h(x) \leq 0\} = 4, \tag{4.10}$$
$$\operatorname{argmin}_{x \in \mathbb{R}^2} \{f(x) | g(x) = 0, h(x) \leq 0\} = \left\{ \begin{bmatrix} 3 \\ 3 \end{bmatrix} \right\} = \{x^\star\}.$$

Quadratic program, continued

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

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

- where we have omitted the constant term in the objective.
- For Problem (4.10), 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**.
- This format can represent AC optimal power flow, which we will formulate in Section 9.1.

Example

$$\min_{x \in \mathbb{R}^3} \{ f(x) | g(x) = 0, h(x) \leq 0 \}, \tag{4.11}$$

- 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 4.1.14 that we can also classify problems on the basis of the notion of **convexity**.

4.1.9.2 Satisfaction of constraints

Definition 4.3 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 $x^\star$. $\square$

Example

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

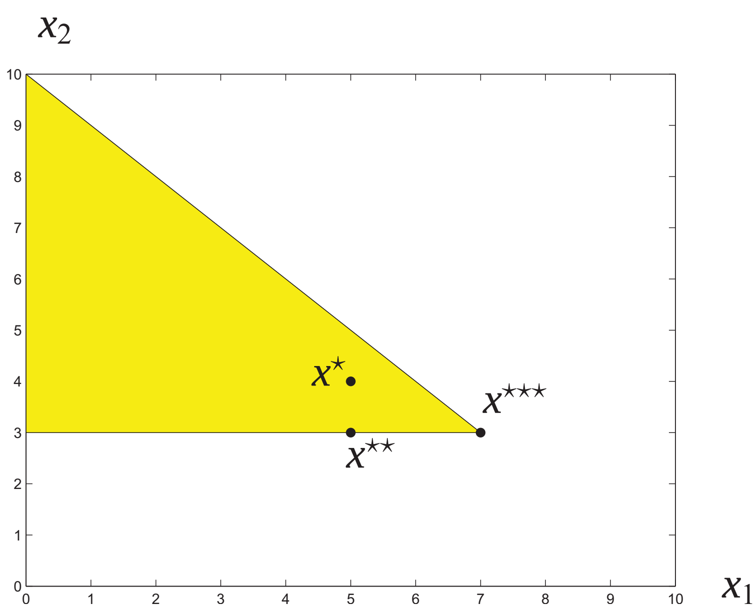

Fig. 4.8. Points $x^\star, x^{\star\star}$, and $x^{\star\star\star}$ that are feasible with respect to inequality constraints. The feasible set is the shaded triangular region for which $x_2 \geq 3$ and $x_1 + x_2 \leq 10$.

Example, continued

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

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

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

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

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

- The constraints $h_1(x) \le 0$ and $h_2(x) \le 0$ 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 $\{x \in \mathbb{R}^2 | h(x) \le \mathbf{0}\}$.

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.

4.1.10 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.
- For larger problems we will use special-purpose software to find the minimum and minimizer:
  - the PowerWorld optimal power flow solver is an example of special-purpose software that is particularly tailored to power systems optimization problems.

4.1.11 Problems without minimum and the infimum

4.1.11.1 Analysis

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

Definition 4.4 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:

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

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. $\square$

4.1.11.2 Example

Unconstrained problem with unbounded objective

$$\forall x \in \mathbb{R}, f(x) = x. \tag{4.12}$$

- 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 $\inf_{x \in \mathbb{R}} f(x) = -\infty$.

4.1.12 Maximization problems and the supremum

$$\max_{x \in \mathbb{S}} f(x) = -\min_{x \in \mathbb{S}}(-f(x)). \tag{4.13}$$

Definition 4.5 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. $\square$

4.1.13 *Solutions of optimization problems*

4.1.13.1 *Local and global minima*

- Recall Problem (4.2) and its minimum $f^\star$:

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

- Sometimes, we call $f^\star$ in Problem (4.2) 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 4.6 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:

$$\exists \varepsilon > 0 \text{ such that } \forall x \in \mathbb{S}, (\|x - x^\star\| < \varepsilon) \Rightarrow (f(x^\star) \leq f(x)). \tag{4.14}$$

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

4.1.13.2 Convex sets

Definition 4.7 Let $\mathbb{S} \subseteq \mathbb{R}^n$. We say that $\mathbb{S}$ is a **convex set** or that $\mathbb{S}$ is **convex** if $\forall x, x' \in \mathbb{S}, \forall t \in [0,1], (1-t)x + tx' \in \mathbb{S}$. $\square$

- A line segment joining any two points in a convex set $\mathbb{S}$ is itself entirely contained in $\mathbb{S}$.

Examples of convex sets

- A line segment is a convex set.

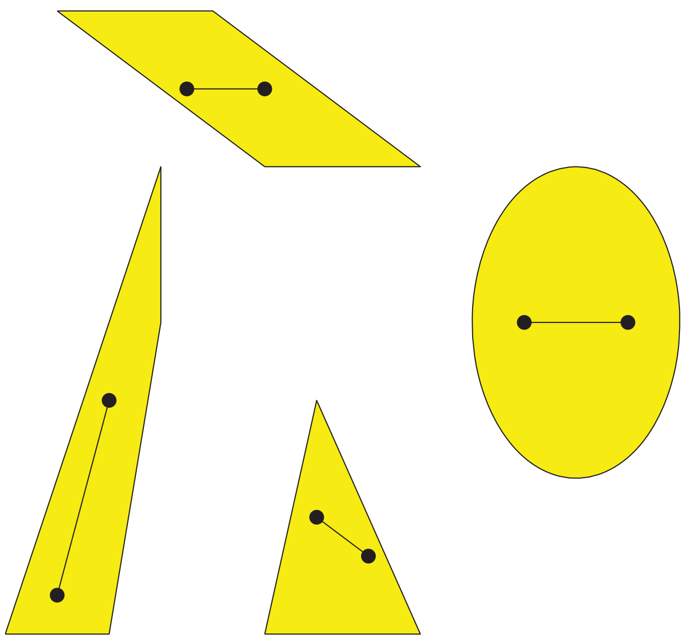

Fig. 4.9. Convex sets with pairs of points joined by line segments.

Examples of non-convex sets

- The union of two non-overlapping line segments is non-convex.
- Non-convex sets can have "indentations."

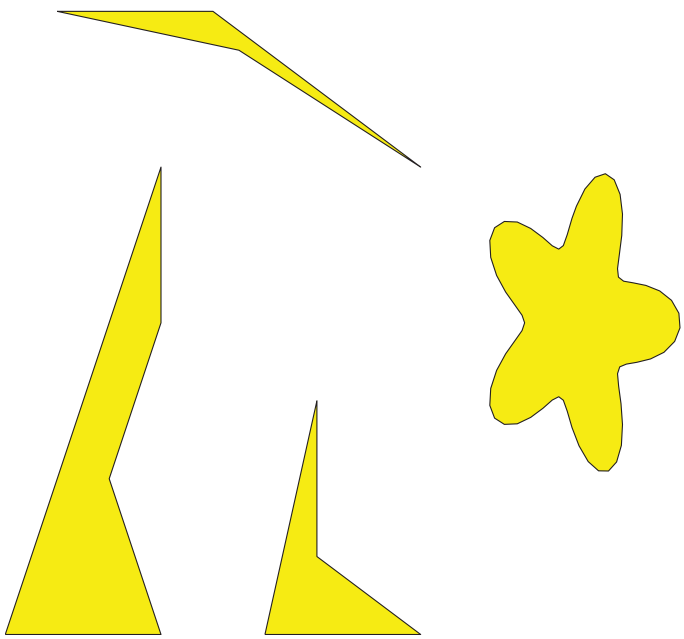

Fig. 4.10. Non-convex sets.

4.1.13.3 Examples of local and global minimizers

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 $\mathbb{S}$.

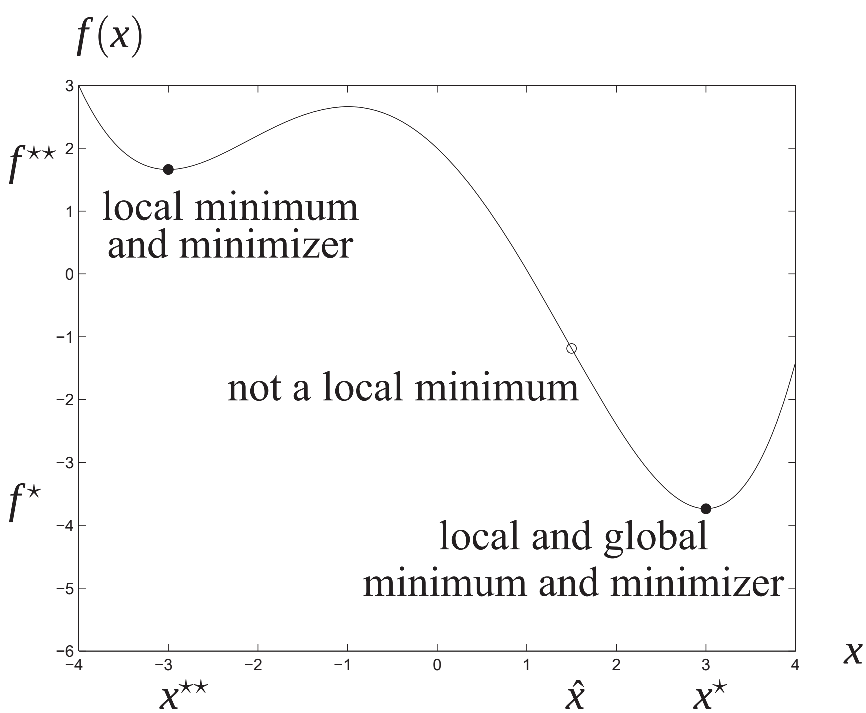

Fig. 4.11. Local minima, $f^\star$ and $f^{\star\star}$, with corresponding local minimizers $x^\star$ and $x^{\star\star}$, over a convex set $\mathbb{S} = \{x \in \mathbb{R} | -4 \leq x \leq 4\}$. The point $x^\star$ is the global minimizer and $f^\star$ the global minimum over $\mathbb{S}$.

Multiple local minimizers over a non-convex set

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

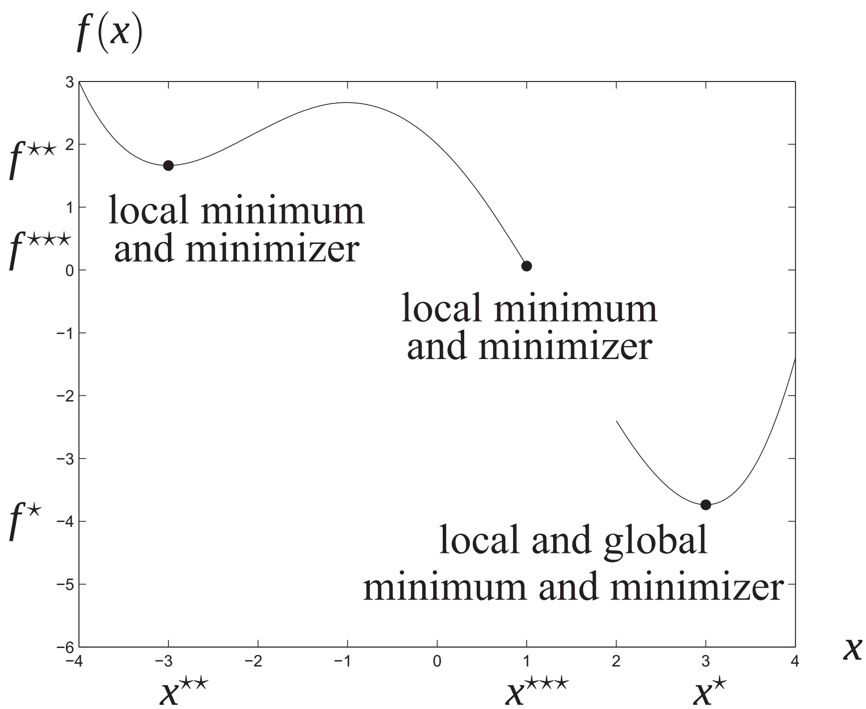

Fig. 4.12. Local and global minima and minimizers of a problem over a non-convex set $\mathbb{P} = \{x \in \mathbb{R} | -4 \le x \le 1 \text{ or } 2 \le x \le 4\}$.

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 $x^{\star\star} \approx \begin{bmatrix} 0.8 \\ -0.7 \end{bmatrix}$.

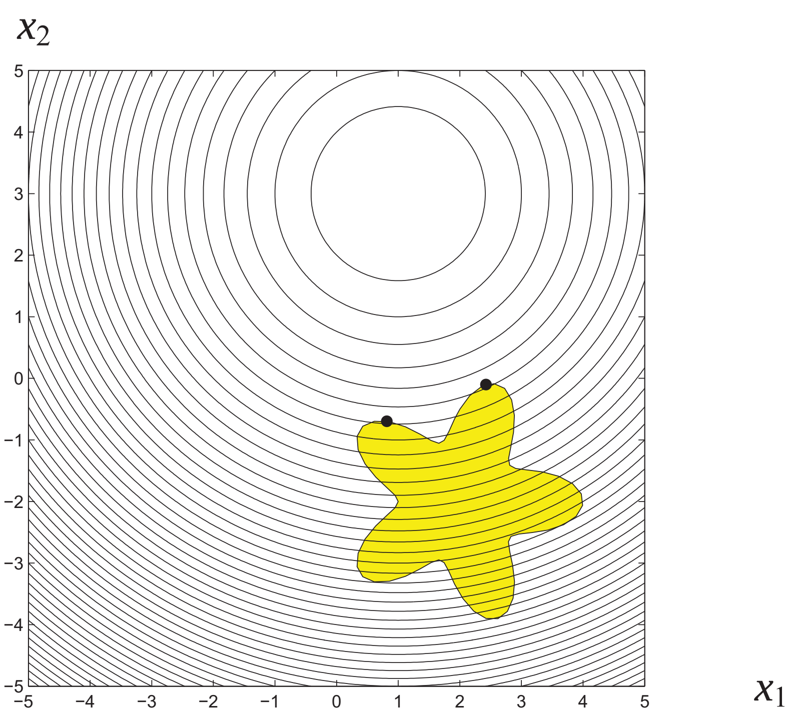

Fig. 4.13. Contour sets of the function defined in (4.4) with feasible set shaded. The two local minimizers are indicated by bullets. The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

4.1.13.4 Discussion

- Iterative improvement algorithms, as typically used in minimizing problems defined in terms of continuous variables, 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.
- For the problem illustrated in Figure 4.13, if an iterative improvement algorithm were started at the point $x^{(0)} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, what would you expect as the result?

4.1.13.5 Strict and non-strict minimizers

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

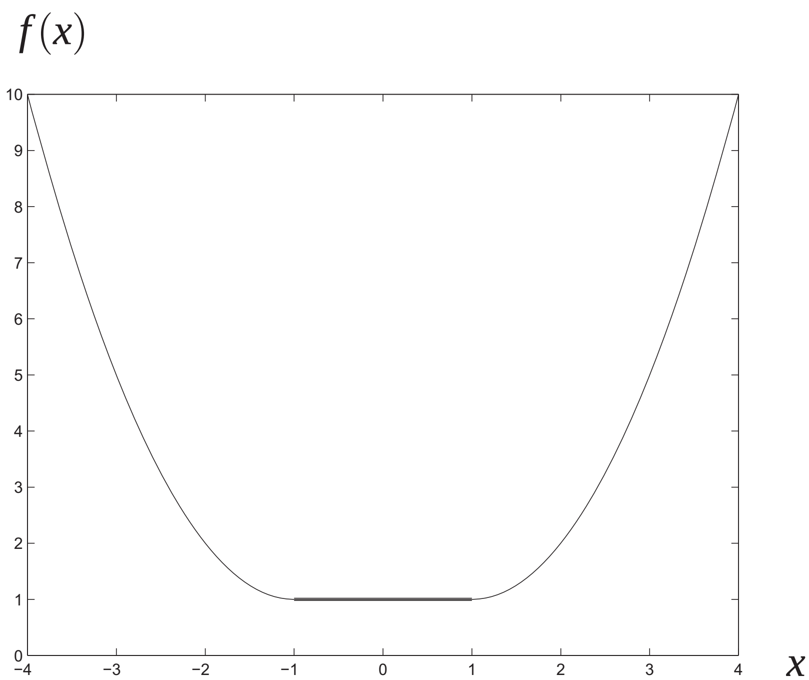

Fig. 4.14. A function with multiple global minimizers. The set of minimizers is indicated by a thick line.

4.1.14 Convex functions

4.1.14.1 Definitions

Definition 4.8 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:

$$\forall x, x' \in \mathbb{S}, \forall t \in [0,1], f([1-t]x + tx') \leq [1-t] f(x) + t f(x'). \tag{4.15}$$

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:

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

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. $\square$

Definitions, continued

- The condition in (4.15) means that linear interpolation of convex f between points on the curve is never below the function values.

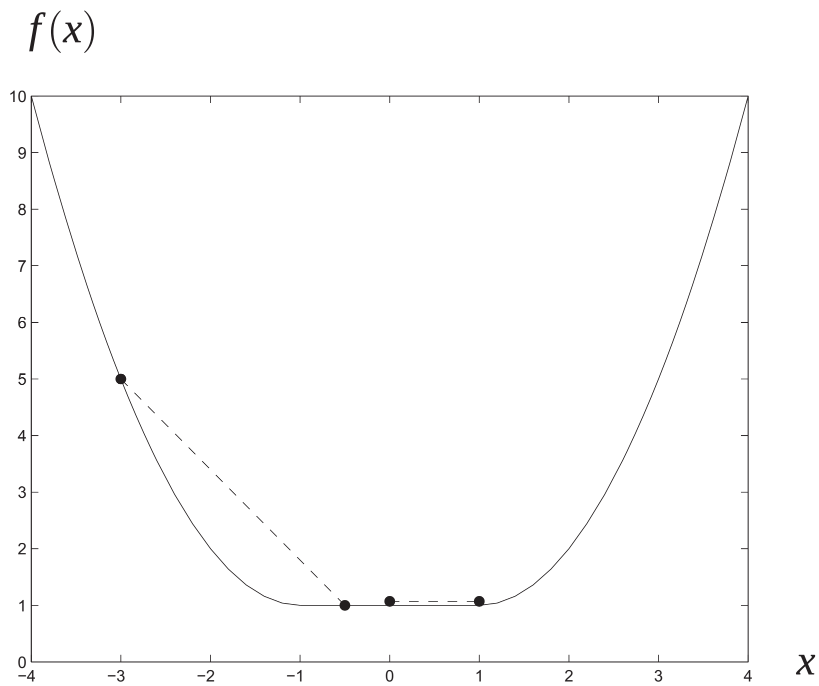

Fig. 4.15. Linear interpolation of a convex function between points never under-estimates the function. (For clarity, the line interpolating f between $x = 0$ and $x = 1$ is drawn slightly above the solid curve: it should be coincident with the solid curve.)

Definitions, continued

Definition 4.9 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 $\mathbb{S}$. $\square$

4.1.14.2 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 4.11 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.

4.1.14.3 Relationship to optimization problems

Theorem 4.1 *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 $\mathbb{S}$.*

$\square$

Definition 4.10 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**. $\square$

- 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.

4.1.14.4 Discussion

- Theorem 4.1 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.
- Iterative improvement algorithms can find the global minima of convex problems.
- Non-convex problems are generally much more difficult to solve.
- We will consider this in more detail in the context of integer problems and will bear this in mind when we formulate electricity market problems as optimization problems.

4.1.14.5 Characterizing convex functions

First derivative

Theorem 4.2 *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:*

$$\forall x, x' \in \mathbb{S}, f(x) \geq f(x') + \nabla f(x')^\dagger (x - x'). \tag{4.16}$$

$\square$

- Recall from Section 2.3.2.1 that the function $\phi : \mathbb{R}^n \to \mathbb{R}$ on the right-hand side of (4.16) 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 (4.16) shows that the first-order Taylor approximation of a convex function never over-estimates the function.

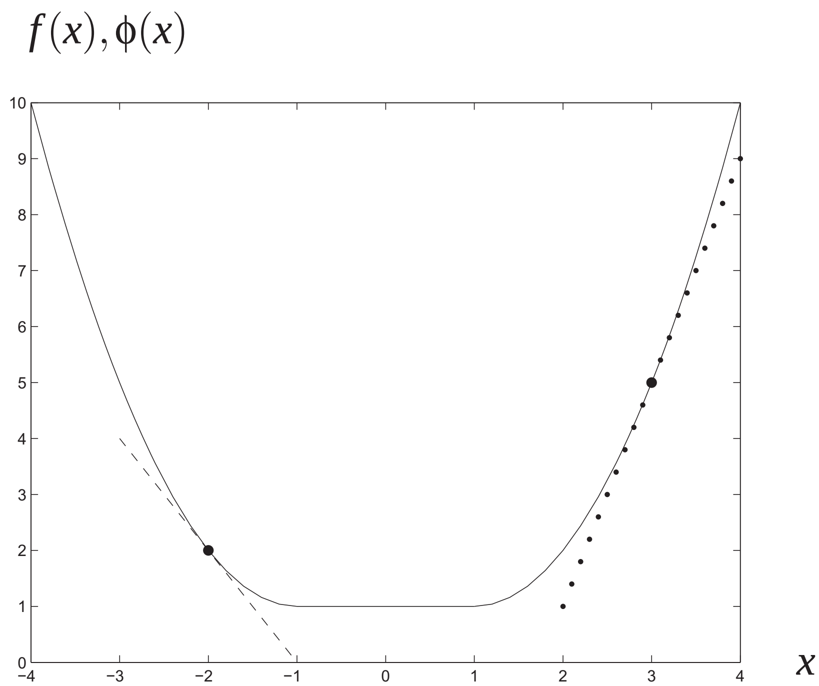

Fig. 4.16. First order Taylor approximation about $x = -2$ (shown dashed) and about $x = 3$ (shown dotted) of a convex function (shown solid).

Second derivative

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

Theorem 4.3 *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 $\mathbb{S}$.* $\Box$

- Which of the following matrices are positive semi-definite?

$$\mathbf{0}, \mathbf{I}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\dagger, \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

- Which of the following matrices are positive definite?

$$\mathbf{0}, \mathbf{I}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

4.1.14.6 *Quadratic functions*

$$\forall x \in \mathbb{R}^n, f(x) = \frac{1}{2}x^\dagger Qx + c^\dagger x, \tag{4.17}$$

- 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 4.3, f is convex.
- If Q is positive definite then, by Theorem 4.3, f is strictly convex.
- If $Q = \mathbf{0}$, so that f is linear, what can you say about the convexity of f?

4.2 Duality

- Taking the **dual** of a constrained 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.
- The constraints in the original problem are said to be **dualized**.
- For reasons that will become clear as we discuss optimality conditions, duality has a close relationship with prices.
- 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:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, h(x) \leq \mathbf{0} \}. \tag{4.18}$$

- 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.

4.2.1 Lagrangian

Definition 4.11 Consider the function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}$ defined by:

$$\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^\dagger g(x) + \mu^\dagger h(x). \quad (4.19)$$

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. $\square$

- 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:

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

4.2.2 Dual function

- Associated with the Lagrangian, we make:

Definition 4.12 Consider the function $\mathcal{D} : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R} \cup \{-\infty\}$ defined by:

$$\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r}, \mathcal{D}(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu). \tag{4.20}$$

The function $\mathcal{D}$ is called the **dual 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 4.9 of a concave function.
- The usefulness of the dual function stems in part from the following:

Theorem 4.4 *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 (4.19), the dual function defined in (4.20), 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.

4.2.3 Dual problem

Theorem 4.5 *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 (4.18). Then:*

$$f(\hat{x}) \geq \mathcal{D}(\lambda, \mu), \tag{4.21}$$

where $\mathcal{D} : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R} \cup \{-\infty\}$ is the dual function defined in (4.20).

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 4.5 enables us to gauge whether we are close to a minimum of Problem (4.18).
- For any value of $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}_+^r$, we know that the minimum of Problem (4.18) is no smaller than $\mathcal{D}(\lambda, \mu)$, which is a lower bound for the problem.
- We can also take the **partial dual** with respect to only some of the constraints leaving the remaining constraints explicit in the definition of the dual function.

Corollary 4.6 *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:*

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

where $\mathbb{E}$ is the effective domain of $\mathcal{D}$. Moreover, if the problem on the left-hand side has a minimum then:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, h(x) \le \mathbf{0} \} \ge \sup_{\left[\begin{smallmatrix}\lambda \\ \mu\end{smallmatrix}\right] \in \mathbb{E}} \{ \mathcal{D}(\lambda, \mu) | \mu \ge \mathbf{0} \}. \tag{4.22}$$

□

Discussion

- This result is called **weak duality**.
- The right-hand side of (4.22) is called the **dual problem** and the constraints in the original problem are said to have been **dualized**.
- In this context, the original problem is sometimes called the **primal problem**.
- By Theorem 4.4, maximizing the dual problem is equivalent to minimizing a convex problem.
- The inequalities in (4.21) and (4.22) 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-hand 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 (4.22) 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 (4.22) can be replaced by equality and the sup and inf can be replaced by max and min, so that the right-hand side of (4.22) equals the minimum of Problem (4.18) and the right-hand side becomes:

$$\max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}}\{\mathcal{D}(\lambda,\mu)|\mu\geq\mathbf{0}\}=\max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}}\left\{\min_{x\in\mathbb{R}^n}\{f(x)+\lambda^\dagger g(x)+\mu^\dagger h(x)\}\middle|\mu\geq\mathbf{0}\right\}, \tag{4.23}$$

- having an inner minimization problem embedded in an outer maximization problem.
- The requirements for these conditions to hold depend on the convexity of the primal problem and on other technical conditions on the functions.
- In the next section, we will consider an example where such conditions happen to hold, and we will discuss sufficient conditions later.
- In the dual problem, the equality and inequality constraints have been transformed into terms in the Lagrangian, which is the objective of the inner minimization problem.

4.2.4 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:

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

- We take the dual with respect to the equality constraint $3-x=0$.
- 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:

$$\begin{aligned}
\max_{\lambda \in \mathbb{E}} \{\mathcal{D}(\lambda)\} &= \max_{\lambda \in \mathbb{R}} \left\{ 3\lambda - \frac{(\lambda)^2}{4} \right\}, \\
&= \max_{\lambda \in \mathbb{R}} \left\{ -\left(\frac{\lambda}{2} - 3\right)^2 + 9 \right\}, \\
&= 9, \text{ with maximizer } \lambda^\star = 6.
\end{aligned}$$

- 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.
- Since $\min_{x \in \mathbb{R}^n} \{f(x) | g(x) = \mathbf{0}\} = \max_{\lambda \in \mathbb{E}} \{\mathcal{D}(\lambda)\}$, there is no duality gap.
- The value $\lambda^\star = 6$ is called the **Lagrange multiplier** for this problem:
  - we will carefully define and generalize the notion of Lagrange multipliers in Section 4.4.1 and subsequently.

4.2.5 Economic interpretation

- We can interpret this example in an economic context relating, for example, to **economic dispatch**.
- Suppose that f is the operating cost of a generator.
- Suppose that $3 - x = 0$ or, equivalently, the demand is 3 and we want to meet the demand with production (or supply) x.
- We consider paying a price π for production x:
  - **revenue** is $\pi \times x$, and
  - production costs are $f(x)$.
- We must model the decision-making process of the generator in response to the prices:
  - revenue minus production costs is called **operating profit**,
  - we model the generator as an **operating profit maximizer**,
  - that is, it seeks to maximize revenue minus production costs, or equivalently minimize production costs minus revenues,
  - operating profit does not include the cost of equipment or any other costs that are not affected by operating decisions.

Economic interpretation, continued

- We claim that setting the price π equal to the Lagrange multiplier $\lambda^\star$ induces a generator trying to maximize its operating profit to meet the demand:

$$\begin{aligned}\text{Operating profit} &= \text{revenue} - \text{production cost},\\ &= (\lambda^\star \times x) - f(x).\end{aligned}$$

- Note that $\mathcal{L}(x,\lambda^\star) = f(x) + \lambda^\star(3-x)$ is minus the operating profit plus a term that is independent of x:

$$\mathcal{L}(x,\lambda^\star) = -\text{Operating profit } + (\lambda^\star \times 3).$$

- To maximize $(\lambda^\star \times x) - f(x) = -\mathcal{L}(x,\lambda^\star) + (\lambda^\star \times 3)$ over values of x, we can equivalently minimize $\mathcal{L}(x,\lambda^\star)$ over values of x.
- The minimizer of $\mathcal{L}(x,\lambda^\star)$ is $x^\star = 3$.
- The minimum (or infimum) of $\mathcal{L}(x,\lambda)$ is $\mathcal{D}(\lambda)$, so $\mathcal{D}(\lambda^\star) = \mathcal{L}(x^\star,\lambda^\star)$.
- In this case, the value of λ that maximizes $\mathcal{D}$ is also the price that induces a profit-maximizing generator to supply $x^\star = 3$, which meets the demand.
- Generalizations of this interpretation will be very important in our discussion of **pricing rules** for electricity markets.

4.3 Continuous unconstrained problems

- We will analyze particular types of optimization problems in detail with the ultimate goal of treating all the significant types of problems that arise in electricity markets, including problems with:
  - equality and inequality constraints, and
  - continuous and discrete variables.
- To build up to the problems we need to treat, we will first consider continuous unconstrained optimization problems of the form:

$$\min_{x \in \mathbb{R}^n} f(x),$$

- where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$.
- In some cases, our objectives will only be piece-wise partially differentiable; however, for convenience here we will assume that the objective is partially differentiable with continuous partial derivatives.
- The extensions for objectives with "kinks" will be discussed as they arise.
- We consider **optimality conditions** that help us to characterize when we have found a minimizer of a problem.

4.3.1 Optimality conditions

Theorem 4.7 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be partially differentiable with continuous partial derivatives. If $x^\star$ is a local minimizer of f then $\nabla f(x^\star) = \mathbf{0}$.* $\square$

- A point that satisfies $\nabla f(x^\star) = \mathbf{0}$ is called a **critical point**.
- A critical point may be a minimizer, a maximizer, or an inflection point of a function.
- With additional information, we can guarantee that a critical point is a minimizer:

Theorem 4.8 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and partially differentiable with continuous partial derivatives on $\mathbb{R}^n$ and let $x^\star \in \mathbb{R}^n$. If $\nabla f(x^\star) = \mathbf{0}$ then $f(x^\star)$ is the global minimum and $x^\star$ is a global minimizer of f.* $\square$

4.3.2 Example

- Consider the objective function defined in (4.4) $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by:

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

- and illustrated in Figures 4.2 and 4.3.
- From Figure 4.3, the minimizer of f is $x^\star = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

Example, continued

- Note that:

$$\begin{aligned} \forall x \in \mathbb{R}^2, \nabla f(x) &= \begin{bmatrix} 2(x_1 - 1) \\ 2(x_2 - 3) \end{bmatrix}, \\ \forall x \in \mathbb{R}^2, \nabla^2 f(x) &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \end{aligned}$$

- which is positive definite.
- Note that $\nabla f(x^\star) = \mathbf{0}$ and, by Theorem 4.3, since $\nabla^2 f$ is positive definite, f is convex.
- Therefore, by Theorem 4.8, $x^\star = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ is a global minimizer of f.

4.4 Continuous equality-constrained problems

- Next, we will consider continuous equality-constrained optimization problems of the form:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b \}, \tag{4.24}$$

- where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$.

4.4.1 Optimality conditions

Theorem 4.9 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. If $x^\star \in \mathbb{R}^n$ is a local minimizer of the problem:*

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

then:

$$\begin{aligned} \exists \lambda^\star \in \mathbb{R}^m \text{ such that } \nabla f(x^\star) + A^\dagger\lambda^\star &= \mathbf{0}, \qquad (4.25)\\ Ax^\star &= b. \qquad (4.26) \end{aligned}$$

$\square$

- A vector $\lambda^\star$ satisfying (4.25), given an $x^\star$ that also satisfies (4.26), is called a vector of **Lagrange multipliers** for the problem.
- The conditions (4.25)–(4.26) are called the **first-order necessary conditions** (or **FONC**) for Problem (4.24).

4.4.2 Example

- Recall the example equality-constrained Problem (4.5):

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

$$\begin{aligned} \text{where: } \forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\ A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\ b &= \begin{bmatrix} 0 \end{bmatrix}. \end{aligned}$$

- The (unique) local minimizer is at $x^\star = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ with minimum $f^\star = 2$.
- We note that:

$$\begin{aligned} \nabla f(x^\star) + A^\dagger [-2] &= \begin{bmatrix} 2 \\ -2 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} [-2], \\ &= \mathbf{0}, \end{aligned}$$

- which is consistent with Theorem 4.9 for $\lambda^\star = [-2]$.

4.4.3 Economic interpretation

- Recall Definition 4.11 of the **Lagrangian**.
- For a problem with objective $f : \mathbb{R}^n \to \mathbb{R}$ and equality constraints $Ax = b$, with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ the Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is defined by:

$$\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \mathcal{L}(x, \lambda) = f(x) + \lambda^\dagger (Ax - b), \qquad (4.27)$$

  where λ is called the vector of **dual variables** for the problem.
- We also define the gradients of $\mathcal{L}$ with respect to x and λ by, respectively, $\nabla_x \mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial x}\right]^\dagger$ and $\nabla_\lambda \mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial \lambda}\right]^\dagger$.
- That is:

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda) &= \nabla f(x) + A^\dagger \lambda, \\ \nabla_\lambda \mathcal{L}(x, \lambda) &= Ax - b. \end{aligned}$$

- We can interpret the first-order necessary conditions (4.25)–(4.26) using the Lagrangian $\mathcal{L}$.

Economic interpretation, continued

- The first-order necessary conditions imply that $x^\star$ is a critical point of the function $\mathcal{L}(\bullet, \lambda^\star)$ that also satisfies the constraints $Ax = b$.
- A minimizer of the equality-constrained problem is also the unconstrained minimizer of $\mathcal{L}(\bullet, \lambda^\star)$:
  - If $\mathcal{L}(\bullet, \lambda^\star)$ is convex then a point $x^\star$ satisfying $\nabla_x \mathcal{L}(x^\star, \lambda^\star) = \mathbf{0}$ will be an unconstrained minimizer of $\mathcal{L}(\bullet, \lambda^\star)$.
  - If we know $\lambda^\star$ then we can solve the equality-constrained problem without explicitly considering the equality constraints!
  - As in the example in Section 4.2.4.
- The vector of Lagrange multipliers $\lambda^\star$ "adjusts" the unconstrained optimality conditions by $A^\dagger \lambda^\star$ to "balance" the minimization of the objective against satisfaction of the constraints.

Economic, interpretation, continued

- As in Section 4.2.5, again interpreting $\mathcal{L}(\bullet, \lambda^\star)$ as minus the profit (plus a constant) to a firm:
  - finding the minimizer of $\mathcal{L}(\bullet, \lambda^\star)$ is equivalent to finding the maximizer of operating profit,
  - the price $\pi = \lambda^\star$ provides the compensation for operating costs incurred by the firm so that unconstrained maximization of operating profits is consistent with minimizing the operating costs subject to the equality constraints.

4.4.4 Example

- Continuing with the previous equality-constrained Problem (4.5), the Lagrangian $\mathcal{L} : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ is defined by:

$$\forall x \in \mathbb{R}^2, \forall \lambda \in \mathbb{R}, \mathcal{L}(x, \lambda) = (x_1 - 1)^2 + (x_2 - 3)^2 + \lambda(x_1 - x_2). \quad (4.28)$$

- Setting the value of the dual variable in the Lagrangian equal to the Lagrange multiplier, $\lambda^\star = [-2]$, we have:

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

- The first-order necessary conditions for minimizing $\mathcal{L}(x, \lambda^\star)$ with respect to x is that:

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda^\star) &= \begin{bmatrix} 2(x_1 - 1) - 2 \\ 2(x_2 - 3) + 2 \end{bmatrix}, \\ &= \mathbf{0}, \end{aligned}$$

- which yields a solution of $x^\star = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

Example, continued

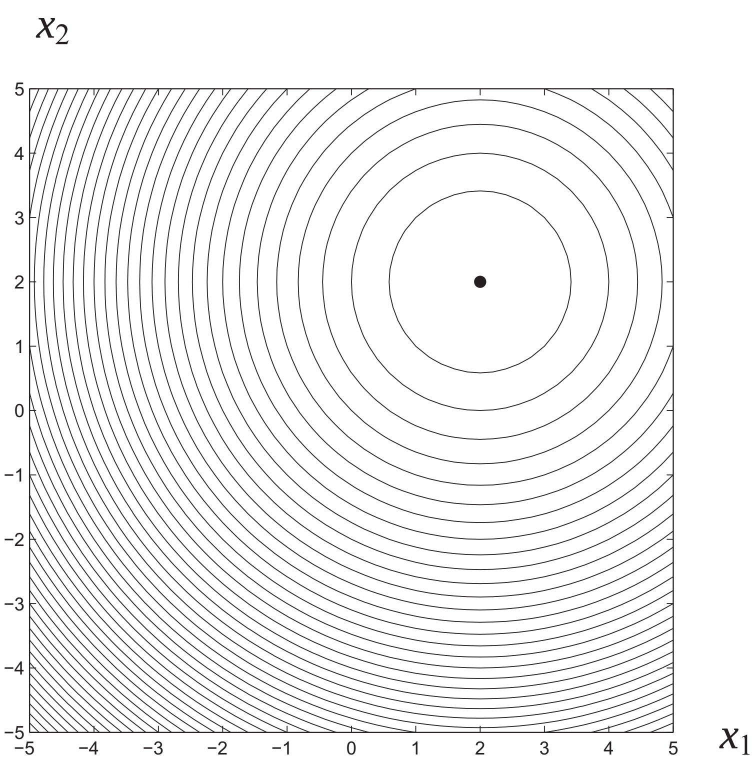

Fig. 4.17. Contour sets for Lagrangian $\mathcal{L}(\bullet, \lambda^\star)$ evaluated at the Lagrange multipliers $\lambda^\star = [-2]$.

Example, continued

- For other values of the dual variables λ not equal to the Lagrange multipliers $\lambda^\star$, the corresponding minimizer of $\mathcal{L}(\bullet, \lambda)$ will differ from the minimizer of Problem (4.5).
- For $\tilde{\lambda} = [-5]$, the contour sets of $\mathcal{L}(\bullet, \tilde{\lambda})$ are illustrated in Figure 4.18.
- The unconstrained minimizer of this function is at $\tilde{x} = \begin{bmatrix} 3.5 \\ 0.5 \end{bmatrix}$, illustrated with a $\circ$ in Figure 4.18, which differs from $x^\star$.
- In the context of our profit interpretation, note that the "wrong" price $\tilde{\lambda}$ will induce the wrong behavior by a profit maximizing firm.

Example, continued

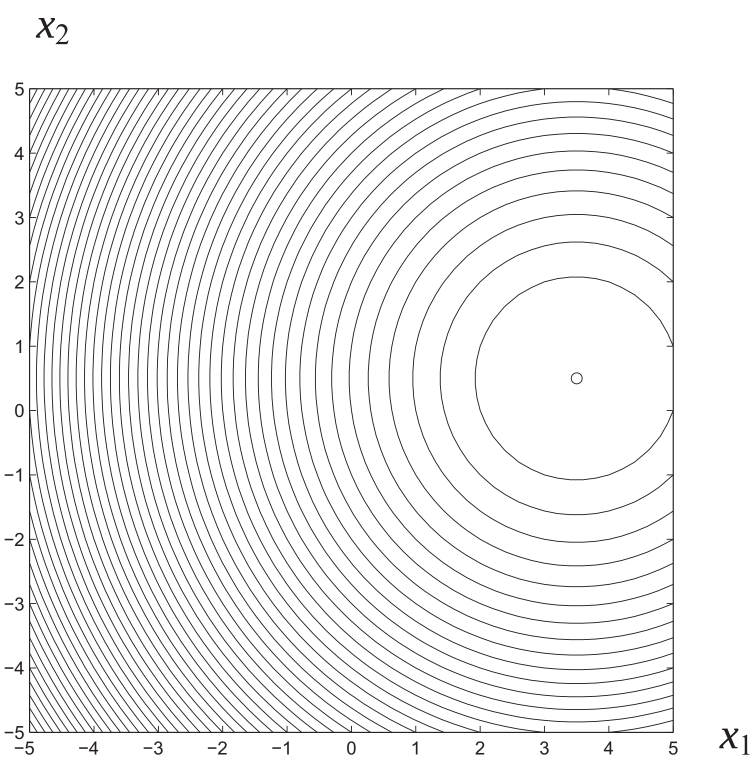

Fig. 4.18. Contour sets for Lagrangian $\mathcal{L}(\bullet, \tilde{\lambda})$ evaluated at value of dual variables $\tilde{\lambda} = [-5]$ not equal to Lagrange multiplers.

4.4.5 Duality

- The discussion in Section 4.4.3 suggests that if we knew the vector of Lagrange multipliers $\lambda^\star$ we could avoid explicit consideration of the equality constraints if f was convex.
- Here we discuss how to characterize the Lagrange multipliers using duality.

4.4.5.1 Dual function

Analysis

- As we discussed in Section 4.2, we can define a dual problem where the role of variables and constraints is partly or fully swapped.
- Recall Definition 4.12 of the **dual function** and **effective domain**.
- For Problem (4.24), the dual function $\mathcal{D} : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$ is defined by:

$$\forall \lambda \in \mathbb{R}^m, \mathcal{D}(\lambda) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda), \tag{4.29}$$

- while the effective domain is:

$$\mathbb{E} = \{\lambda \in \mathbb{R}^m | \mathcal{D}(\lambda) > -\infty\},$$

- so that the restriction of $\mathcal{D}$ to $\mathbb{E}$ is a function $\mathcal{D} : \mathbb{E} \to \mathbb{R}$.

Example

$$\begin{aligned}\forall x \in \mathbb{R}^2, \forall \lambda \in \mathbb{R}, \mathcal{L}(x,\lambda) &= (x_1-1)^2+(x_2-3)^2+\lambda(x_1-x_2),\\ \forall \lambda \in \mathbb{R}, \mathcal{D}(\lambda) &= \inf_{x\in\mathbb{R}^2} \mathcal{L}(x,\lambda),\\ &= \inf_{x\in\mathbb{R}^2} \{(x_1-1)^2+(x_2-3)^2+\lambda(x_1-x_2)\}.\end{aligned}$$

- $\mathcal{L}(\bullet,\lambda)$ is partially differentiable with continuous partial derivatives and is strictly convex.
- By Corollary 4.8 the first-order necessary conditions are sufficient for global optimality:

$$\begin{aligned}\nabla_x \mathcal{L}(x,\lambda) &= \begin{bmatrix} 2(x_1-1)+\lambda \\ 2(x_2-3)-\lambda \end{bmatrix},\\ &= \mathbf{0}.\end{aligned}$$

Example, continued

- For any given $\lambda \in \mathbb{R}$, the unique solution is $x^{(\lambda)} = \begin{bmatrix} 1-\lambda/2 \\ 3+\lambda/2 \end{bmatrix}$.

$$\begin{aligned} \forall \lambda \in \mathbb{R}, \mathcal{D}(\lambda) &= \left(1-\frac{\lambda}{2}-1\right)^2 + \left(3+\frac{\lambda}{2}-3\right)^2 + \lambda\left(1-\frac{\lambda}{2}-3-\frac{\lambda}{2}\right), \\ &= -\frac{(\lambda)^2}{2} - 2\lambda. \end{aligned} \tag{4.30}$$

4.4.5.2 Dual problem

Analysis

- As we illustrated in Section 4.2.4, under certain conditions, Lagrange multipliers $\lambda^\star$ can be found as the maximizer, over the dual variables λ, of the **dual problem**:

$$\max_{\lambda\in\mathbb{E}} \mathcal{D}(\lambda). \tag{4.31}$$

- Problem (4.31) is called the **dual problem** to Problem (4.24).
- Problem (4.24) is called the **primal problem**.
- Moreover, under certain conditions, the corresponding minimizer $x^{(\lambda^\star)}$ of the inner problem $\min_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda^\star)$ satisfies the equality constraints.

Theorem 4.10 *Suppose that* $f : \mathbb{R}^n \to \mathbb{R}$ *is convex and partially differentiable with continuous partial derivatives,* $A \in \mathbb{R}^{m\times n}$*, and* $b \in \mathbb{R}^m$*. Consider primal problem, Problem (4.24):*

$$\min_{x\in\mathbb{R}^n}\{f(x)|Ax=b\}.$$

Also, consider the dual problem, Problem (4.31). If the primal problem possesses a minimum then the dual problem possesses a maximum and the optima are equal. That is:

$$\min_{x\in\mathbb{R}^n}\{f(x)|Ax=b\} = \max_{\lambda\in\mathbb{E}} \mathcal{D}(\lambda). \tag{4.32}$$

□

- Recall from Theorem 4.4 that the effective domain $\mathbb{E}$ of the dual function is a convex set and that the dual function is concave on $\mathbb{E}$.
- This facilitates finding a solution of the dual problem.

Example

- Continuing with the previous equality-constrained Problem (4.5), the dual function $\mathcal{D}$ was specified in (4.30).
- The dual function is partially differentiable with continuous partial derivatives on the whole of $\mathbb{R}$.
- Moreover, since the dual function is concave, the first-order necessary conditions to maximize $\mathcal{D}$ are also sufficient.
- Partially differentiating $\mathcal{D}$ we obtain:

$$\nabla\mathcal{D}(\lambda) = [-\lambda - 2].$$

- Moreover, $\nabla\mathcal{D}(\lambda) = [0]$ for $\lambda^\star = [-2]$.
- Also, $\mathcal{D}(\lambda^\star) = 2$, which is equal to the minimum of Problem (4.5) and $x^{(\lambda^\star)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, which is the minimizer of Problem (4.5).

4.4.5.3 Separable objective

Analysis

- Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is additively separable, so that:

$$\forall x \in \mathbb{R}^n, f(x) = \sum_{k=1}^{n} f_k(x_k),$$

- where $f_k : \mathbb{R} \to \mathbb{R}, k = 1, \ldots, n$.
- We consider the dual.

Analysis, continued

$$
\begin{aligned}
\forall \lambda \in \mathbb{E}, \mathcal{D}(\lambda) &= \inf_{x \in \mathbb{R}^n} \mathcal{L}(x,\lambda), \\
&= \min_{x \in \mathbb{R}^n} \mathcal{L}(x,\lambda), \text{ assuming that the minimum exists,} \\
&= \min_{x \in \mathbb{R}^n} f(x) + \lambda^\dagger (Ax - b), \text{ by definition of } \mathcal{L}, \\
&= \min_{x \in \mathbb{R}^n} \left\{ \sum_{k=1}^{n} f_k(x_k) + \lambda^\dagger \left(\sum_{k=1}^{n} A_k x_k - b \right) \right\}, \\
&\quad \text{where } A_k \text{ is the } k\text{-th column of } A, \\
&= \min_{x \in \mathbb{R}^n} \left\{ \sum_{k=1}^{n} \left(f_k(x_k) + \lambda^\dagger A_k x_k \right) \right\} - \lambda^\dagger b, \\
&= \sum_{k=1}^{n} \min_{x_k \in \mathbb{R}} \{ f_k(x_k) + \lambda^\dagger A_k x_k \} - \lambda^\dagger b. \qquad (4.33)
\end{aligned}
$$

Analysis, continued

- For each fixed $\lambda \in \mathbb{R}^m$, the dual function $\mathcal{D}(\lambda)$ is the sum of:
  - a constant $(-\lambda^\dagger b)$, and
  - n one-dimensional optimization "sub-problems" that can each be evaluated independently.
- We have **decomposed** the problem by exploiting the separability of the objective.
- We can think of each of the decomposed problems as corresponding to operating profit maximization for individual firms given a price specified by the value of the dual variables:
  - again, as in Sections 4.2.5 and 4.4.3, the price $\pi = \lambda^\star$ provides the compensation for operating costs incurred by each firm so that unconstrained maximization of operating profits for each firm is consistent with minimizing the overall operating costs subject to the equality constraints,
  - typically, the equality constraints will be supply–demand balance constraints.

Example

- Continuing with the previous equality-constrained Problem (4.5), note that the objective is separable.
- The dual function is:

$$\begin{aligned} \forall \lambda \in \mathbb{R}, \mathcal{D}(\lambda) &= \min_{x \in \mathbb{R}^2} \mathcal{L}(x, \lambda), \\ &= \min_{x_1 \in \mathbb{R}} \{(x_1 - 1)^2 + \lambda x_1\} + \min_{x_2 \in \mathbb{R}} \{(x_2 - 3)^2 - \lambda x_2\}. \end{aligned} \tag{4.34}$$

- Each of the two convex sub-problems can be solved separately and the result is the same as obtained previously, with the same value of Lagrange multiplier $\lambda^\star$.
- If the the sub-problems correspond to operating profit maximization for each firm:
  - the price $\pi = \lambda^\star$ provides the compensation for operating costs incurred by each firm so that unconstrained maximization of operating profits for each firm is consistent with minimizing the overall operating costs subject to the equality constraints.

4.4.6 Sensitivity analysis

Theorem 4.11 *Consider Problem (4.24), a perturbation vector $\gamma \in \mathbb{R}^m$, and a perturbed version of Problem (4.24) defined by:*

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b - \gamma \}. \tag{4.35}$$

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice partially differentiable with continuous second partial derivatives, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$, with the rows of A linearly independent. Let $x^\star \in \mathbb{R}^n$ and $\lambda^\star \in \mathbb{R}^m$ satisfy:

$$\begin{aligned} \nabla f(x^\star) + A^\dagger \lambda^\star &= \mathbf{0}, \\ Ax^\star &= b, \\ ((A\Delta x = \mathbf{0}) \text{ and } (\Delta x \neq \mathbf{0})) &\Rightarrow (\Delta x^\dagger \nabla^2 f(x^\star) \Delta x > 0). \end{aligned}$$

Consider Problem (4.35). For values of γ in a neighborhood of the base-case value of the parameters $\gamma = \mathbf{0}$, there is a local minimum and corresponding local minimizer and Lagrange multipliers for Problem (4.35). Moreover, the local minimum, local minimizer, and Lagrange multipliers are partially differentiable with respect to γ and *have continuous partial derivatives in this neighborhood. The sensitivity of the local minimum to γ, evaluated at the base-case $\gamma = \mathbf{0}$, is equal to $[\lambda^\star]^\dagger$. If f is convex then the minimizers and minima are global.* $\square$

4.4.7 Discussion

- The sufficient conditions for the sensitivity theorem are not always satisfied by the problems we study.
- Nevertheless, the sensitivity analysis can give us powerful economic insights.

Discussion, continued

- If we assume that the minimizer and minimum are well-defined functions of γ and that they are partially differentiable with respect to γ, then the following argument explains why the sensitivity is given by the value of the Lagrange multipliers.
- Consider Problem (4.35), a perturbation γ, and the corresponding change $\Delta x^\star$ in the minimizer of the perturbed problem.
- The change in the minimum is:

$$\begin{aligned} f(x^\star + \Delta x^\star) - f(x^\star) &\approx \nabla f(x^\star)^\dagger \Delta x^\star, \text{ with equality as } \Delta x^\star \rightarrow \mathbf{0}, \\ &= -[\lambda^\star]^\dagger A \Delta x^\star, \text{ by the first-order} \\ &\quad\ \text{necessary condition } \nabla f(x^\star) + A^\dagger \lambda^\star = \mathbf{0}, \\ &= [\lambda^\star]^\dagger \gamma, \end{aligned}$$

- since $A(x^\star + \Delta x^\star) = b - \gamma$, so that $-A\Delta x^\star = \gamma$.
- But this is true for any such perturbation γ. In the limit as $\gamma \rightarrow \mathbf{0}$, the change in the minimum approaches $[\lambda^\star]^\dagger \gamma$.

Discussion, continued

- We can interpret the Lagrange multipliers as the sensitivity of the minimum to changes in γ.
- In many problems, the specification of constraints represents some judgment about the availability of resources.
- Then we can use the Lagrange multipliers to help in trading off the change in the optimal objective against the cost of the purchase of additional resources.
- In particular, if the equality constraint represents supply–demand balance then the Lagrange multiplier provides information about the marginal cost of meeting additional demand.

4.4.8 Example

- Consider the equality-constrained Problem (4.5) from Section 4.1.8:

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

$$\begin{aligned}\forall x\in\mathbb{R}^2, f(x) &= (x_1-1)^2+(x_2-3)^2,\\ A &= \begin{bmatrix}1 & -1\end{bmatrix},\\ b &= \begin{bmatrix}0\end{bmatrix}.\end{aligned}$$

- Suppose that the equality constraints changed from $Ax=b$ to $Ax=b-\gamma$.
- Then, if γ is small enough, the minimum of the perturbed problem differs from the minimum of the original problem by approximately $[\lambda^\star]^\dagger\gamma=(-2)\gamma$.

4.5 Continuous linear inequality-constrained problems

- Next, we consider inequality-constrained optimization problems of the form:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, Cx \leq d \}, \tag{4.36}$$

- where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{r \times n}$, and $d \in \mathbb{R}^r$ are constants.
- We call the constraints $Cx \leq d$ **linear inequality constraints**.

4.5.1 Optimality conditions

Theorem 4.12 *Suppose that* $f : \mathbb{R}^n \rightarrow \mathbb{R}$ *is partially differentiable with continuous partial derivatives,* $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, C \in \mathbb{R}^{r \times n}, d \in \mathbb{R}^r$. *Consider Problem (4.36):*

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b, Cx \leq d \},$$

and a point $x^\star \in \mathbb{R}^n$. *If* $x^\star$ *is a local minimizer of Problem (4.36) then:*

$$\begin{aligned} \exists \lambda^\star \in \mathbb{R}^m, \exists \mu^\star \in \mathbb{R}^r \text{ such that: } \nabla f(x^\star) + A^\dagger \lambda^\star + C^\dagger \mu^\star &= \mathbf{0}; \\ M^\star (Cx^\star - d) &= \mathbf{0}; \\ Ax^\star &= b; \\ Cx^\star &\leq d; \text{ and} \\ \mu^\star &\geq \mathbf{0}, \end{aligned} \quad (4.37)$$

where $M^\star = \operatorname{diag}\{\mu_\ell^\star\} \in \mathbb{R}^{r \times r}$.

The vectors $\lambda^\star$ and $\mu^\star$ satisfying the conditions (4.37) are called the vectors of Lagrange multipliers for the constraints $Ax = b$ and $Cx \leq d$, respectively. The conditions that $M^\star(Cx^\star - d) = \mathbf{0}$ are called the **complementary slackness conditions***. They say that, for each ℓ, either the ℓ-th inequality constraint is binding or the ℓ-th Lagrange multiplier is equal to zero (or both).* $\square$

4.5.2 Example

- Recall the example quadratic program, Problem (4.10):

$$\min_{x \in \mathbb{R}^2} \{ f(x) | Ax = b, Cx \leq d \}.$$

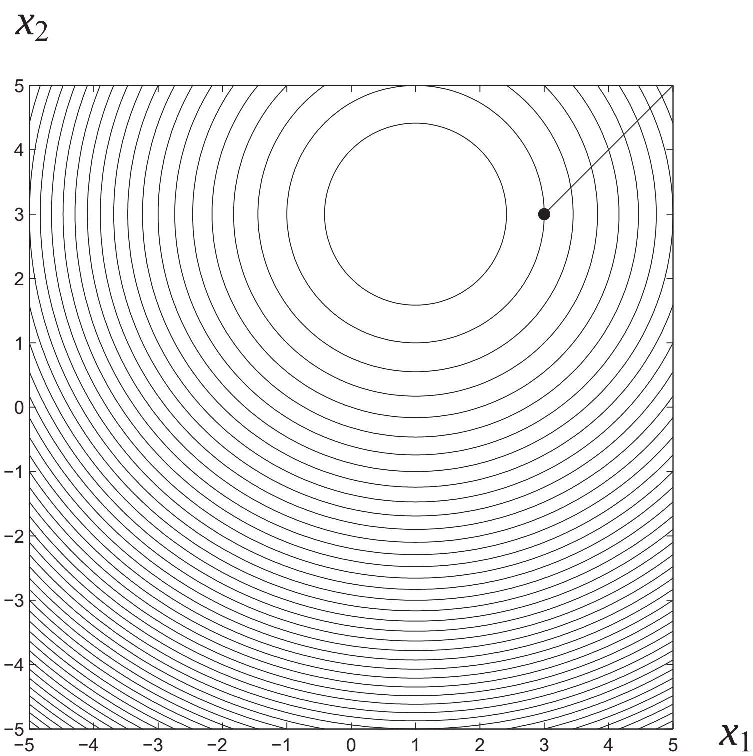

Fig. 4.19. Contour sets of objective function and feasible set for Problem (4.10). The heights of the contours decrease towards the point $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$. The feasible set is the "half-line" starting at the point $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$, which is also the minimizer and is illustrated with a $\bullet$.

Example, continued

- The objective and constraints are specified by:

$$\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\
A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\
b &= \begin{bmatrix} 0 \end{bmatrix}, \\
C &= \begin{bmatrix} 0 & -1 \end{bmatrix}, \\
d &= \begin{bmatrix} -3 \end{bmatrix}.
\end{aligned}$$

- Figure 4.19 shows that the solution of this problem is $x^\star = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$.
- We claim that $x^\star = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ together with $\lambda^\star = [-4]$ and $\mu^\star = [4]$ satisfy (4.37) for Problem (4.10).

Example, continued

$$\begin{aligned}
\forall x \in \mathbb{R}^2, \nabla f(x) &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} x + \begin{bmatrix} -2 \\ -6 \end{bmatrix}, \\
\nabla f(x^\star) + A^\dagger \lambda^\star + C^\dagger \mu^\star & \\
&= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} + \begin{bmatrix} -2 \\ -6 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} [-4] + \begin{bmatrix} 0 \\ -1 \end{bmatrix} [4], \\
&= \mathbf{0}; \\
\mu^\star (Cx^\star - d) &= [4] \left(\begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} - [-3] \right), \\
&= [0]; \\
Ax^\star &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \\
&= [0], \\
&= b;
\end{aligned}$$

Example, continued

$$
\begin{aligned}
Cx^\star &= \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \\
&= [-3], \\
&\le [-3], \\
&= d; \text{ and} \\
\mu^\star &= [4], \\
&\ge [0].
\end{aligned}
$$

4.5.3 Discussion

- The Lagrange multipliers again adjust the unconstrained optimality conditions to balance the constraints against the objective:
  - since the inequality constraints only need to be "enforced" in one direction, the Lagrange multipliers on the inequality constraints are restricted in sign.
- We will again refer to the equality and inequality constraints in (4.37) as *the* first-order necessary conditions, although we recognize that the first-order necessary conditions also include, strictly speaking, the other items in the hypothesis of Theorem 4.12.
- These conditions are called the **Kuhn–Tucker** (KT) or the **Karush–Kuhn–Tucker** (KKT) conditions and a point satisfying the conditions is called a **KKT point**.

4.5.4 Lagrangian

- Recall Definition 4.11 of the **Lagrangian**.
- For Problem (4.36) the Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}$ is defined by:

$$\forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m, \forall \mu \in \mathbb{R}^r, \mathcal{L}(x,\lambda,\mu) = f(x) + \lambda^\dagger(Ax-b) + \mu^\dagger(Cx-d).$$

- As in the equality-constrained case, define the gradients of $\mathcal{L}$ with respect to x, λ, and μ by, respectively, $\nabla_x \mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial x}\right]^\dagger$, $\nabla_\lambda \mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial \lambda}\right]^\dagger$, and $\nabla_\mu \mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial \mu}\right]^\dagger$.
- Evaluating the gradients with respect to x, λ, and μ, we have:

$$\begin{aligned}
\nabla_x \mathcal{L}(x,\lambda,\mu) &= \nabla f(x) + A^\dagger \lambda + C^\dagger \mu, \\
\nabla_\lambda \mathcal{L}(x,\lambda,\mu) &= Ax - b, \\
\nabla_\mu \mathcal{L}(x,\lambda,\mu) &= Cx - d.
\end{aligned}$$

- Setting the first two of these expressions equal to zero reproduces some of the first-order necessary conditions for the problem.

Lagrangian, continued

- As with equality-constrained problems, the Lagrangian provides a convenient way to remember the optimality conditions.
- However, unlike the equality-constrained case, in order to recover the first-order necessary conditions for Problem (4.36) we have to:
  - add the complementary slackness conditions; that is, $M^\star(Cx^\star - d) = \mathbf{0}$,
  - add the non-negativity constraints on μ, that is, $\mu \geq \mathbf{0}$, and
  - interpret the third expression on the previous slide as corresponding to inequality constraints; that is, $Cx \leq d$.
- If the hypotheses of Theorem 4.12 are satisfied and, additionally, f is convex then $x^\star$ is a global minimizer of $\mathcal{L}(\bullet, \lambda^\star, \mu^\star)$, where $\lambda^\star$ and $\mu^\star$ are the Lagrange multipliers.

Lagrangian, continued

- Paralleling earlier discussion, interpreting $\mathcal{L}(\bullet, \lambda^\star, \mu^\star)$ as minus the profit (plus a constant) to a firm:
  - finding the minimizer of $\mathcal{L}(\bullet, \lambda^\star, \mu^\star)$ is equivalent to finding the maximizer of profit,
  - the prices $\lambda^\star$ and $\mu^\star$ provide the compensation for operating costs incurred by the firm so that unconstrained maximization of profits is consistent with minimizing the operating costs subject to the equality and inequality constraints.

4.5.5 Convex problems

Theorem 4.13 *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, C \in \mathbb{R}^{r\times n}, d \in \mathbb{R}^r$. Consider Problem (4.36):*

$$\min_{x\in\mathbb{R}^n}\{f(x)|Ax = b, Cx \leq d\},$$

and points $x^\star \in \mathbb{R}^n$, $\lambda^\star \in \mathbb{R}^m$, and $\mu^\star \in \mathbb{R}^r$. Let $M^\star = \mathrm{diag}\{\mu^\star_\ell\}$. Suppose that:

(i) *f is convex on $\{x \in \mathbb{R}^n|Ax = b, Cx \leq d\}$,*
(ii) *$\nabla f(x^\star) + A^\dagger\lambda^\star + C^\dagger\mu^\star = \mathbf{0}$,*
(iii) *$M^\star(Cx^\star - d) = \mathbf{0}$,*
(iv) *$Ax^\star = b$ and $Cx^\star \leq d$, and*
(v) *$\mu^\star \geq \mathbf{0}$.*

Then $x^\star$ is a global minimizer of Problem (4.36). $\square$

- In addition to the first-order necessary conditions, the first-order sufficient conditions require that f is convex on the feasible set.

4.5.6 Example

- Again consider Problem (4.10) from Sections 4.1.9 and 4.5.2.
- In Section 4.5.2, we observed that $x^\star = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, $\lambda^\star = [-4]$, and $\mu^\star = [4]$ satisfy the first-order necessary conditions for this problem.
- Moreover, f is twice continuously differentiable with continuous partial derivatives and the Hessian is positive definite.
- Therefore, f is convex and $x^\star$ is the global minimizer of the problem.

4.5.7 Duality

- As we discussed in Section 4.2 and as in the discussion of linear equality constraints in Section 4.4.5, we can define a dual problem where the role of variables and constraints is partly or fully swapped.
- We again recall some of the discussion in Section 4.2 in the following sections.

4.5.7.1 *Dual function*

- We have observed in Section 4.5.4 that if f is convex then $x^\star$ is a global minimizer of $\mathcal{L}(\bullet, \lambda^\star, \mu^\star)$.
- Recall Definition 4.12 of the **dual function** and **effective domain**.
- For Problem (4.36), the dual function $\mathcal{D} : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R} \cup \{-\infty\}$ is defined by:

$$\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^{m+r}, \mathcal{D}(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu). \tag{4.38}$$

- The effective domain of $\mathcal{D}$ is:

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

- Recall that by Theorem 4.4, $\mathbb{E}$ is convex and $\mathcal{D}$ is concave on $\mathbb{E}$.

Example

- We continue with Problem (4.10).
- The problem is:

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

- where:

$$\begin{aligned} \forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\ A &= [1 \quad -1], \\ b &= [0], \\ C &= [0 \quad -1], \\ d &= [-3]. \end{aligned}$$

- The Lagrangian $\mathcal{L} : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ for this problem is defined by:

$$\begin{aligned} &\forall x \in \mathbb{R}^2, \forall \lambda \in \mathbb{R}, \forall \mu \in \mathbb{R}, \\ \mathcal{L}(x, \lambda, \mu) &= f(x) + \lambda^\dagger (Ax - b) + \mu^\dagger (Cx - d), \\ &= (x_1 - 1)^2 + (x_2 - 3)^2 + \lambda [1 \quad -1] x + \mu ([0 \quad -1] x + 3). \end{aligned}$$

Example, continued

- For any given λ and μ, the Lagrangian $\mathcal{L}(\bullet,\lambda,\mu)$ is strictly convex.
- By Corollary 4.8, the first-order necessary conditions $\nabla_x\mathcal{L}(x,\lambda,\mu)=0$ are sufficient for minimizing $\mathcal{L}(\bullet,\lambda,\mu)$.
- Moreover, a minimizer exists, so that the inf in the definition of $\mathcal{D}$ can be replaced by min.
- Furthermore, there is a unique minimizer $x^{(\lambda,\mu)}$ corresponding to each value of λ and μ:

$$
\begin{aligned}
&\forall x\in\mathbb{R}^2,\forall\lambda\in\mathbb{R},\forall\mu\in\mathbb{R},\\
&\qquad\qquad \nabla_x\mathcal{L}(x,\lambda,\mu) = \nabla f(x)+A^\dagger\lambda+C^\dagger\mu,\\
&\qquad\qquad\qquad = \begin{bmatrix}2&0\\0&2\end{bmatrix}x+\begin{bmatrix}-2\\-6\end{bmatrix}+\begin{bmatrix}1\\-1\end{bmatrix}\lambda+\begin{bmatrix}0\\-1\end{bmatrix}\mu,\\
&\forall\lambda\in\mathbb{R},\forall\mu\in\mathbb{R},x^{(\lambda,\mu)} = -\begin{bmatrix}2&0\\0&2\end{bmatrix}^{-1}\left[\begin{bmatrix}-2\\-6\end{bmatrix}+\begin{bmatrix}1\\-1\end{bmatrix}\lambda+\begin{bmatrix}0\\-1\end{bmatrix}\mu\right],\\
&\qquad\qquad\qquad = \begin{bmatrix}1\\3\end{bmatrix}+\begin{bmatrix}-0.5\\0.5\end{bmatrix}\lambda+\begin{bmatrix}0\\0.5\end{bmatrix}\mu. \qquad (4.39)
\end{aligned}
$$

Example, continued

- Consequently, the effective domain is $\mathbb{E} = \mathbb{R} \times \mathbb{R}$ and the dual function $\mathcal{D} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is given by:

$$
\begin{aligned}
\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^2, \mathcal{D}(\lambda, \mu) &= \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu), \\
&= \mathcal{L}(x^{(\lambda,\mu)}, \lambda, \mu), \text{ since } x^{(\lambda,\mu)} \text{ minimizes } \mathcal{L}(\bullet, \lambda, \mu), \\
&= (x_1^{(\lambda,\mu)} - 1)^2 + (x_2^{(\lambda,\mu)} - 3)^2 \\
&\quad + \lambda [1 \ \ -1] x^{(\lambda,\mu)} + \mu \left([0 \ \ -1] x^{(\lambda,\mu)} + 3 \right), \\
&= -\frac{1}{2}(\lambda)^2 - \frac{1}{4}(\mu)^2 - 2\lambda - \frac{1}{2}\mu\lambda,
\end{aligned}
$$

- on substituting from (4.39) for $x^{(\lambda,\mu)}$.

4.5.7.2 *Dual problem*

Analysis

- As in the equality-constrained case, if the objective is convex on $\mathbb{R}^n$ then the minimum of Problem (4.36) is equal to $\mathcal{D}(\lambda^\star, \mu^\star)$, where $\lambda^\star$ and $\mu^\star$ are the Lagrange multipliers that satisfy the necessary conditions for Problem (4.36).
- As in the equality-constrained case, under certain conditions, the Lagrange multipliers can be found as the maximizer of the **dual problem**:

$$\max_{\left[\begin{smallmatrix}\lambda\\ \mu\end{smallmatrix}\right]\in\mathbb{E}} \{\mathcal{D}(\lambda, \mu) | \mu \geq \mathbf{0}\}, \tag{4.40}$$

- where $\mathcal{D} : \mathbb{E} \to \mathbb{R}$ is the dual function defined in (4.38).
- Again, Problem (4.36) is called the **primal problem** to distinguish it from Problem (4.40).

Example

- Continuing with the dual of Problem (4.10), the effective domain is $\mathbb{E} = \mathbb{R} \times \mathbb{R}$ and the dual function $\mathcal{D} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is:

$$\forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^2, \mathcal{D}(\lambda, \mu) = -\frac{1}{2}(\lambda)^2 - \frac{1}{4}(\mu)^2 - 2\lambda - \frac{1}{2}\mu\lambda,$$

- with unique minimizer of the Lagrangian specified by (4.39).
- The dual function is twice partially differentiable with continuous second partial derivatives.
- In particular,

$$\begin{aligned} \forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^2, \nabla\mathcal{D}(\lambda, \mu) &= \begin{bmatrix} -2 - \lambda - \mu/2 \\ -\lambda/2 - \mu/2 \end{bmatrix}, \\ \forall \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \in \mathbb{R}^2, \nabla^2\mathcal{D}(\lambda, \mu) &= \begin{bmatrix} -1 & -0.5 \\ -0.5 & -0.5 \end{bmatrix}. \end{aligned}$$

- We claim that $\begin{bmatrix} \lambda^\star \\ \mu^\star \end{bmatrix} = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$ maximizes the dual function over $\mu \geq [0]$.

Example, continued

- In particular $\nabla\mathcal{D}(\lambda^\star, \mu^\star) = \mathbf{0}$, $\mu^\star > [0]$, and $\nabla^2\mathcal{D}$ is negative definite.
- Consequently, $\begin{bmatrix} \lambda^\star \\ \mu^\star \end{bmatrix}$ is the unique maximizer of dual Problem (4.40).

4.5.8 Sensitivity analysis

Theorem 4.14 *Consider perturbations* $\gamma \in \mathbb{R}^m$ *and* $\eta \in \mathbb{R}^r$ *and the problem:*

$$\min_{x \in \mathbb{R}^n} \{ f(x) | Ax = b - \gamma, Cx \leq d - \eta \}. \tag{4.41}$$

Suppose that the function $f : \mathbb{R}^n \to \mathbb{R}$ *is twice partially differentiable with continuous second partial derivatives. Suppose that* $x^\star \in \mathbb{R}^n$ *is a local minimizer of Problem (4.41) for the base-case values* $\gamma = \mathbf{0}$ *and* $\eta = \mathbf{0}$*, with associated Lagrange multipliers* $\lambda^\star$ *and* $\mu^\star$*. Moreover, suppose that the matrix* $\hat{A}$ *has linearly independent rows, where* $\hat{A}$ *is the matrix with rows consisting of:*

- *the* m *rows of* A*, and*
- *those rows* C_ℓ *of* C *for which* $\ell \in \mathbb{A}(x^\star)$*.*

Furthermore, suppose that there are no inequality constraints that are binding at the base-case solution with corresponding values of Lagrange multipliers zero and that:

$$\left((\hat{A}\Delta x = \mathbf{0}) \text{ and } (\Delta x \neq \mathbf{0}) \right) \Rightarrow (\Delta x^\dagger \nabla^2 f(x^\star) \Delta x > 0).$$

Then, for values of γ *and* η *in a neighborhood of the base-case value of the parameters* $\gamma = \mathbf{0}$ *and* $\eta = \mathbf{0}$*, there is a local minimum and corresponding local minimizer and Lagrange multipliers for Problem (4.41). Moreover, the local minimum, local minimizer, and Lagrange multipliers are partially differentiable with respect to* γ *and* η *and have continuous partial derivatives. The sensitivities of the local minimum to* γ *and* η*, evaluated at the base-case* $\gamma = \mathbf{0}$ *and* $\eta = \mathbf{0}$*, are equal to* $[\lambda^\star]^\dagger$ *and* $[\mu^\star]^\dagger$*, respectively.* $\square$

4.5.9 Discussion

- The Lagrange multipliers yield the sensitivity of the objective to the right-hand side of the equality constraints and inequality constraints.
- Again, the sufficient conditions for the sensitivity theorem are not always satisfied by the problems we study.
- Theorem 4.14 does not apply directly to linear programming problems; however, sensitivity analysis can also be applied to linear programming and, as with linear programming in general, the linearity of both objective and constraints leads to various special cases.
- Again, the Lagrange multipliers associated with the constraints can give us powerful economic insights.
- Suppose that there are some inequality constraints that are binding at the base-case solution having corresponding values of Lagrange multipliers zero. Why is the value of the Lagrange multiplier not a reliable indicator of the sensitivity of the minimum to changes in the corresponding right-hand side?

4.5.10 Example

- Consider Problem (4.10) from Sections 4.1.9 and 4.5.2, which has objective $f : \mathbb{R}^2 \to \mathbb{R}$ and constraints $Ax = b$ and $Cx \leq d$ defined by:

$$\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\
A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\
b &= \begin{bmatrix} 0 \end{bmatrix}, \\
C &= \begin{bmatrix} 0 & -1 \end{bmatrix}, \\
d &= \begin{bmatrix} -3 \end{bmatrix}.
\end{aligned}$$

- Note that the binding constraint at the base-case solution has non-zero Lagrange multiplier.
- The matrix:

$$\hat{A} = \begin{bmatrix} A \\ C \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix},$$

- has linearly independent rows and $\nabla^2 f$ is positive definite, so that $(\Delta x \neq \mathbf{0}) \Rightarrow (\Delta x^\dagger \nabla^2 f(x^\star) \Delta x > 0)$.

Example, continued

- Suppose that the inequality constraint was changed to $Cx \leq d - \eta$.
- If η is small enough, then by Theorem 4.14 the minimum of the perturbed problem differs from the minimum of the original problem by approximately $[\mu^\star]^\dagger \eta$.

4.6 Continuous non-linear inequality-constrained problems

- The final type of continuous problem we will consider is:

$$\min_{x \in \mathbb{R}^n} \{ f(x) | g(x) = \mathbf{0}, h(x) \leq \mathbf{0} \}, \tag{4.42}$$

- where $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^r$ are non-linear.

4.6.1 Regular point

Definition 4.13 Let $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^r$. Then we say that $x^\star$ is a **regular point** of the constraints $g(x) = \mathbf{0}$ and $h(x) \leq \mathbf{0}$ if:

(i) $g(x^\star) = \mathbf{0}$ and $h(x^\star) \leq \mathbf{0}$,

(ii) g and h are both partially differentiable with continuous partial derivatives, and

(iii) the matrix $\hat{A}$ has linearly independent rows, where $\hat{A}$ is the matrix with rows consisting of:

- the m rows of the Jacobian $J(x^\star)$ of g evaluated at $x^\star$, and
- those rows $K_\ell(x^\star)$ of the Jacobian K of h evaluated at $x^\star$ for which $\ell \in \mathbb{A}(x^\star)$.

The matrix $\hat{A}$ consists of the rows of $J(x^\star)$ together with those rows of $K(x^\star)$ that correspond to the active constraints. If there are no equality constraints then the matrix $\hat{A}$ consists of the rows of $K(x^\star)$ corresponding to active constraints. If there are no binding inequality constraints then $\hat{A} = J(x^\star)$. If there are no equality constraints and no binding inequality constraints then the matrix $\hat{A}$ has no rows and, by definition, it has linearly independent rows.

$\square$

4.6.2 Optimality conditions

Theorem 4.15 *Suppose that the functions* $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, *and* $h : \mathbb{R}^n \to \mathbb{R}^r$ *are partially differentiable with continuous partial derivatives. Let* $J : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ *and* $K : \mathbb{R}^n \to \mathbb{R}^{r\times n}$ *be the Jacobians of* g *and* h*, respectively. Consider Problem (4.42):*

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

Suppose that $x^\star \in \mathbb{R}^n$ *is a regular point of the constraints* $g(x) = \mathbf{0}$ *and* $h(x) \leq \mathbf{0}$.

If $x^\star$ is a local minimizer of Problem (4.42) then:

$$
\begin{aligned}
\exists \lambda^\star \in \mathbb{R}^m, \exists \mu^\star \in \mathbb{R}^r \text{ such that: } \nabla f(x^\star) + J(x^\star)^\dagger \lambda^\star + K(x^\star)^\dagger \mu^\star &= \mathbf{0}; \\
M^\star h(x^\star) &= \mathbf{0}; \\
g(x^\star) &= \mathbf{0}; \\
h(x^\star) &\leq \mathbf{0}; \text{ and} \\
\mu^\star &\geq \mathbf{0},
\end{aligned}
\tag{4.43}
$$

where $M^\star = \operatorname{diag}\{\mu_\ell^\star\} \in \mathbb{R}^{r \times r}$. The vectors $\lambda^\star$ and $\mu^\star$ satisfying the conditions (4.43) are called the vectors of Lagrange multipliers for the constraints $g(x) = \mathbf{0}$ and $h(x) \leq \mathbf{0}$, respectively. The conditions that $M^\star h(x^\star) = \mathbf{0}$ are called the **complementary slackness conditions**. *They say that, for each ℓ, either the ℓ-th inequality constraint is binding or the ℓ-th Lagrange multiplier is equal to zero (or both).* □

- As previously, we refer to the equality and inequality constraints in (4.43) as the **first-order necessary conditions** (or **FONC**) or the **Karush–Kuhn–Tucker** conditions.

4.6.3 Lagrangian

- Recall Definition 4.11 of the **Lagrangian**.
- Analogously to the discussion in Section 4.5.4, by defining the Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}$ by:

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

- we can again reproduce some of the first-order necessary conditions as:

$$\begin{aligned} \nabla_x \mathcal{L}(x^\star, \lambda^\star, \mu^\star) &= \mathbf{0}, \\ \nabla_\lambda \mathcal{L}(x^\star, \lambda^\star, \mu^\star) &= \mathbf{0}, \\ \nabla_\mu \mathcal{L}(x^\star, \lambda^\star, \mu^\star) &\leq \mathbf{0}. \end{aligned}$$

4.6.4 Example

- Recall the example non-linear program, Problem (4.11), from Section 4.1.9:

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

- 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:

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

- We claim that $x^\star=\begin{bmatrix}0.5\\1.5\\\pi/6\end{bmatrix}$, $\lambda^\star=\begin{bmatrix}6\\1\end{bmatrix}$, and $\mu^\star=[5]$ satisfy the first-order necessary conditions in Theorem 4.15.

Example, continued

- First, $x^\star$ is feasible.

$$
\begin{aligned}
\forall x \in \mathbb{R}^3, \nabla f(x) &= \begin{bmatrix} 2x_1 \\ 4x_2 \\ 0 \end{bmatrix}, \\
\forall x \in \mathbb{R}^3, J(x) &= \begin{bmatrix} 0 & -1 & -\cos(x_3) \\ -1 & 0 & \cos(x_3) \end{bmatrix}, \\
J(x^\star) &= \begin{bmatrix} 0 & -1 & -\cos(\pi/6) \\ -1 & 0 & \cos(\pi/6) \end{bmatrix}, \\
\forall x \in \mathbb{R}^3, K(x) &= \begin{bmatrix} 0 & 0 & \cos(x_3) \end{bmatrix}, \\
K(x^\star) &= \begin{bmatrix} 0 & 0 & \cos(\pi/6) \end{bmatrix}.
\end{aligned}
$$

- Note that $\hat{A} = \begin{bmatrix} J(x^\star) \\ K(x^\star) \end{bmatrix}$ has linearly independent rows so that $x^\star$ is a regular point of the constraints.

Example, continued

$$
\begin{aligned}
\nabla f(x^\star) + J(x^\star)^\dagger \lambda^\star + K(x^\star)^\dagger \mu^\star & \\
&= \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ -1 & 0 \\ -\cos(\pi/6) & \cos(\pi/6) \end{bmatrix} \begin{bmatrix} 6 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \cos(\pi/6) \end{bmatrix} 5, \\
&= \mathbf{0}; \\
\mu^\star h(x^\star) &= [5] \times [0], \\
&= [0]; \\
g(x^\star) &= \mathbf{0}; \\
h(x^\star) &= [0], \\
&\leq [0]; \text{ and} \\
\mu^\star &= [5], \\
&\geq [0].
\end{aligned}
$$

- That is, $x^\star, \lambda^\star$, and $\mu^\star$ satisfy the first-order necessary conditions.

4.6.5 Sensitivity

- We can also develop sensitivity analysis.
- Again, the Lagrange multipliers provide information about sensitivity to changes in the constraints.

4.7 Integer problems

- In some formulations, the entries of the decision vector must be **integer-valued**:
  - the decision to have a generator on or off in **unit commitment** is binary-valued,
  - combined-cycle generating units typically have several discrete operating modes, such as: off; one gas turbine operating; one gas turbine and one steam turbine operating; two gas turbines operating; two gas turbines and one steam turbine operating.
- We write $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ for the set of integers.
- An **integer programming** problem seeks the minimum and minimizer over choices of a decision variable that lies in some subset of $\mathbb{Z}^n$.
- To emphasize that the variables are no longer continuous, we will use the symbol z for decision vectors with entries that are integer-valued.

4.7.1 Example

- Suppose that $\mathbb{S} \subseteq \mathbb{Z}^2$ is the set of points z such that $z_1 \in \{0, 1\}$ and $z_2 \in \{0, 1\}$.
- Why is this set non-convex?

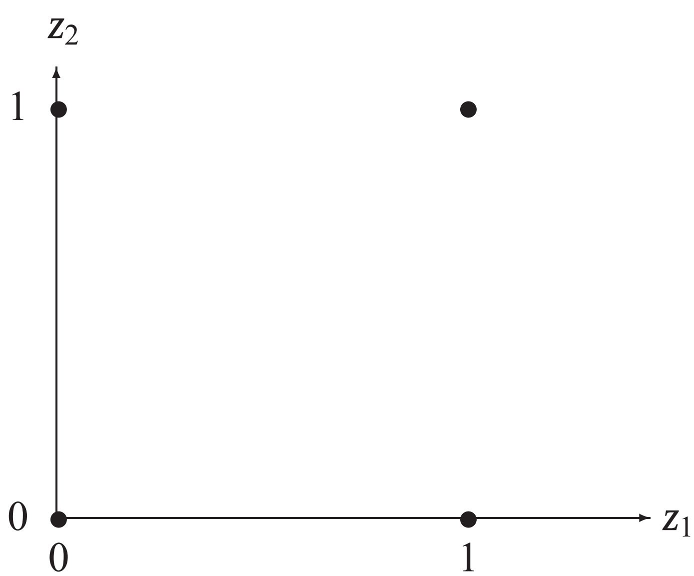

Fig. 4.20. Example feasible set $\mathbb{S}$ for integer program.

4.7.2 Non-convexity of feasible set

- Integer programming problems generally have non-convex feasible sets:
  - The feasible set in the example is non-convex since a line drawn between any two points in the feasible set does not entirely lie in the set.
- Because of the non-convexity of the feasible set, iterative improvement algorithms are usually insufficient to solve integer programming problems:
  - General-purpose algorithms for solving integer programming problems can be extremely computationally intensive.
  - Some particular integer programming problems can be solved efficiently.

4.7.3 Types of problems

- As with optimization problems involving continuous variables, we can consider integer problems with feasible sets that are defined in terms of:
  - **equality constraints**, and
  - **inequality constraints**.
- Commercial software for integer programming is available for **integer linear programs**:

$$\min_{z\in\mathbb{Z}^n}\{c^\dagger z|Az=b,Cz\leq d\},$$

- and **integer quadratic programs**:

$$\min_{z\in\mathbb{Z}^n}\left\{\frac{1}{2}x^\dagger Qx+c^\dagger z\,\middle|\,Az=b,Cz\leq d\right\}.$$

4.7.4 Duality

- Because of the non-convex feasible set, there is usually a duality gap between primal and dual formulations of integer programming problems:
  - In our profit maximization interpretation, this typically means that prices associated with the dual variables are *insufficient* to induce a profit maximizer to behave consistently with minimizing the costs subject to the constraints.
  - In the context of electricity markets that include unit commitment, such as **day-ahead markets**, this means that we need more than prices on energy supply–demand balance to induce generators to be committed consistent with minimizing the costs.
  - We will see that **side payments** are typically used in such electricity markets to induce behavior that is consistent with minimizing overall costs.

4.8 Mixed-integer problems

- In many problems, only some of the entries of the decision vector must be integer-valued, while the others are continuous:
  - the decision to have a generator on or off in **unit commitment** is binary-valued,
  - the production level of the generator is continuous-valued.
- A **mixed-integer programming** problem seeks the minimum and minimizer over choices of decision variables such that some entries have integer values and some have continuous values.

4.8.1 Example

- Suppose that $\mathbb{P} \subseteq \mathbb{R}^2$ is the set of points $\begin{bmatrix} z \\ x \end{bmatrix}$ such that $z \in \{0, 1\}$ and $2z \leq x \leq 4z$.
- Why is this set non-convex?

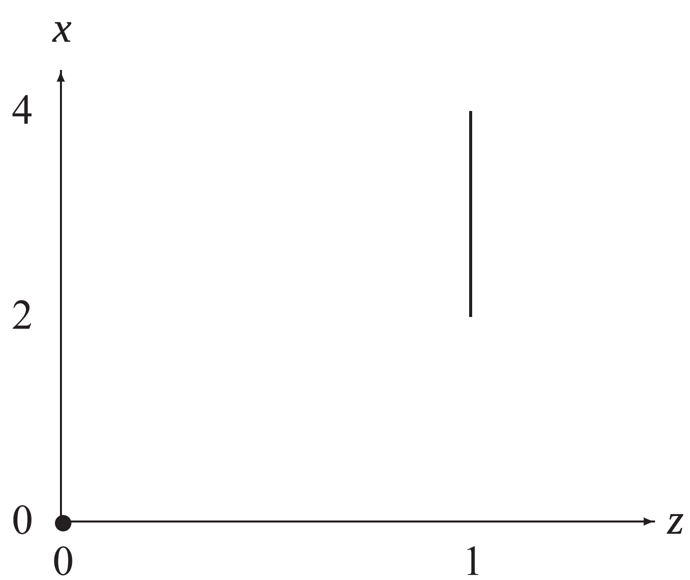

Fig. 4.21. Example feasible set $\mathbb{P}$ for mixed-integer program.

4.8.2 Discussion

- Again, because of the non-convexity of the feasible set, general purpose algorithms are very computationally intensive.
- Commercial software for mixed-integer programming is available for **mixed-integer linear programs** and **mixed-integer quadratic programs**.
- As with integer programming problems, there is usually a duality gap with mixed-integer programming problems.

4.8.3 Example of duality gap

- Consider the problem $\min_{\begin{bmatrix} z \\ x \end{bmatrix} \in \mathbb{P}} \left\{ f\left(\begin{bmatrix} z \\ x \end{bmatrix}\right) \middle| g\left(\begin{bmatrix} z \\ x \end{bmatrix}\right) = 0 \right\}$, where $\mathbb{P} \subseteq \mathbb{R}^2$, $f : \mathbb{R}^2 \to \mathbb{R}$, and $g : \mathbb{R}^2 \to \mathbb{R}$ are defined by:

$$\mathbb{P} = \left\{ \begin{bmatrix} z \\ x \end{bmatrix} \in \mathbb{R}^2 \middle| z \in \{0,1\}, 2z \leq x \leq 4z \right\},$$

so that $\mathbb{P}$ is the earlier example feasible set,

$$\forall x \in \mathbb{P}, f\left(\begin{bmatrix} z \\ x \end{bmatrix}\right) = \begin{cases} 0, & \text{if } z = 0 \text{ and } x = 0, \\ & \text{(with the generator “off”)}, \\ 4 + x, & \text{if } z = 1 \text{ and } 2 \leq x \leq 4, \\ & \text{(with the generator“on” and producing } x\text{)}, \end{cases}$$

$$\forall x \in \mathbb{R}, g\left(\begin{bmatrix} z \\ x \end{bmatrix}\right) = 3 - x, \text{ so if } g\left(\begin{bmatrix} z \\ x \end{bmatrix}\right) = 0 \text{ then supply } x \text{ equals demand } 3.$$

- Note that the generator has two variables associated with its operation:
  - a “unit commitment” variable z, and
  - a “production” variable x.

Example of duality gap, continued

- To meed demand of 3, the generator must be on, so that $z^\star = 1$, and the generator must have production $x^\star = 3$.
- That is, the only feasible point, and therefore the minimizer of this problem, is $z^\star = 1$ and $x^\star = 3$.
- The minimum is $f\left(\begin{bmatrix} z^\star \\ x^\star \end{bmatrix}\right) = 4 + x^\star = 7$.

Example of duality gap, continued

- We consider the dual function $\mathcal{D} : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ with respect to the equality constraint, defined by:

$$\begin{aligned}\forall \lambda \in \mathbb{R}, \mathcal{D}(\lambda) &= \inf_{\begin{bmatrix} z \\ x \end{bmatrix} \in \mathbb{P}} \mathcal{L}\left(\begin{bmatrix} z \\ x \end{bmatrix}, \lambda\right), \\ &= \inf_{\begin{bmatrix} z \\ x \end{bmatrix} \in \mathbb{P}} \left\{ f\left(\begin{bmatrix} z \\ x \end{bmatrix}\right) + \lambda(3 - x) \right\}.\end{aligned}$$

- To minimize the Lagrangian over $\begin{bmatrix} z \\ x \end{bmatrix} \in \mathbb{P}$, we will need to consider the values of λ.
- First note that:

$$\begin{aligned} f\left(\begin{bmatrix} z \\ x \end{bmatrix}\right) + \lambda(3-x) &= \begin{cases} 3\lambda, & \text{if } z = 0 \text{ and } x = 0, \\ 3\lambda + 4 + (1-\lambda)x, & \text{if } z = 1 \text{ and } 2 \leq x \leq 4, \end{cases} \\ &= \begin{cases} 3\lambda, & \text{if } z = 0 \text{ and } x = 0, \\ 3\lambda + 4 - (\lambda - 1)x, & \text{if } z = 1 \text{ and } 2 \leq x \leq 4. \end{cases} \end{aligned}$$

Example of duality gap, continued

- To minimize $\mathcal{L}\left(\begin{bmatrix} z \\ x \end{bmatrix}, \lambda\right) = f\left(\begin{bmatrix} z \\ x \end{bmatrix}\right) + \lambda(3-x)$, we must compare 3λ to values of $3\lambda + 4 - (\lambda - 1)x$ with $2 \leq x \leq 4$.
- We consider various cases for λ.

$\lambda \leq 1$

$$\begin{aligned} 3\lambda &< 3\lambda + 4, \\ &\leq 3\lambda + 4 + (1-\lambda)x, \text{ for } 2 \leq x \leq 4. \end{aligned}$$

- So, the Lagrangian is minimized for $z = 0, x = 0$.
- $\mathcal{D}(\lambda) = 3\lambda$.

$1 < \lambda < 2$

- Then $(\lambda - 1)x < 4$ for $2 \leq x \leq 4$.

$$3\lambda < 3\lambda + 4 - (\lambda - 1)x, \text{ for } 2 \leq x \leq 4.$$

- So, the Lagrangian is again minimized for $z = 0, x = 0$.
- $\mathcal{D}(\lambda) = 3\lambda$.

Example of duality gap, continued

$\lambda = 2$

- Then $3\lambda < 3\lambda + 4 - (\lambda - 1)x$ for $2 \leq x < 4$.
- Also, $3\lambda = 3\lambda + 4 - (\lambda - 1)x$ for $x = 4$.
- So, the Lagrangian has two minimizers:

  $z = 0, x = 0$, and
  $z = 1, x = 4$.
- $\mathcal{D}(\lambda) = 3\lambda$.

$\lambda > 2$

$$3\lambda > 3\lambda + 4 - (\lambda - 1)x, \text{ for } x = 4.$$

- Moreover, the right-hand side decreases with increasing x, so it is minimized over $2 \leq x \leq 4$ by $x = 4$.
- So, the Lagrangian is minimized for $z = 1, x = 4$.
- $\mathcal{D}(\lambda) = 3\lambda + 4 - (\lambda - 1)4 = 8 - \lambda$.

Example of duality gap, continued

- The figure shows that the maximum of the dual occurs at $\lambda^\star = 2$ with $\mathcal{D}(\lambda^\star) = 6$.
- However, the corresponding value of x does not meet demand.

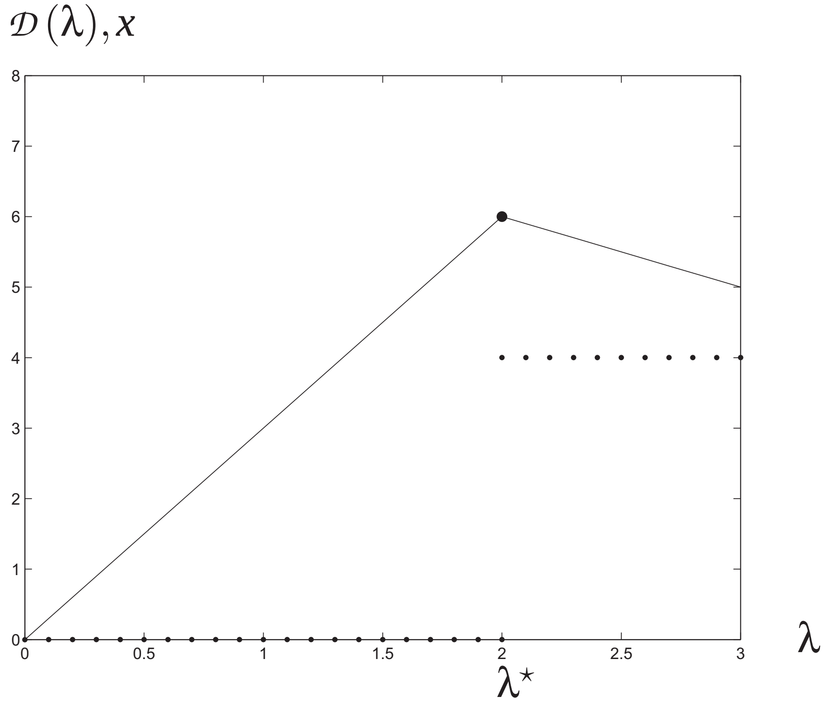

Fig. 4.22. Dual function $\mathcal{D}(\lambda)$ (shown solid) and the corresponding value of x (shown dotted) versus λ for example mixed-integer problem.

4.8.4 Economic interpretation of duality gap

- In this problem there is a duality gap, since:

$$f\left(\begin{bmatrix} z^\star \\ x^\star \end{bmatrix}\right) = 7 > 6 = \mathcal{D}(\lambda^\star).$$

- No price λ for production will equate supply to the demand of 3:
  - For $\lambda < 2$, a profit maximizing firm will produce nothing.
  - For $\lambda = 2$, a profit maximizing firm is indifferent between producing nothing and producing 4 units. It prefers these alternatives to producing at any other level.
  - For $\lambda > 2$, a profit maximizing firm will want to produce 4 units, exceeding demand.
- Whenever there is a duality gap, there are no prices on the corresponding dualized constraints that will induce profit maximizing firms to satisfy the constraints:
  - **side payments** that are separate from the prices on the dualized constraints will be used as part of the pricing rule to induce behavior that is consistent with minimizing overall costs.

4.9 Summary

- In this chapter we have defined optimization problems.
- We illustrated particular types of problems with elementary examples.
- We defined the notion of convexity.
- We defined local and global and strict and non-strict minima and minimizers of optimization problems.
- Continuous, integer and mixed-integer problems were defined.
- Duality and optimality conditions for continuous problems were presented.
- Integer and mixed-integer problems were defined.
- Implications of dualizing non-convex problems was explored.

This chapter is based, in part, on Chapters 2, 10, 13, 17, and 19 of *Applied Optimization: Formulation and Algorithms for Engineering Systems,* Cambridge University Press 2006.

Homework exercises

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

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

(i) Sketch the contour sets $\mathbb{C}_f(\tilde{f})$ for $\tilde{f} = 0, 1, 2, 3$.

(ii) Sketch on the same graph the set of points satisfying $g(x) = 0$ where $g : \mathbb{R}^2 \to \mathbb{R}$ is defined by:

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

(iii) Use your sketch to find the minimum $f^\star$ and the minimizer $x^\star$ of $\min_{x \in \mathbb{R}^2}\{f(x)|g(x) = 0\}$.

(iv) Find a value of the Lagrange multiplier $\lambda^\star$ that satisfies the first-order necessary conditions in Theorem 4.9. (Hint: Theorem 4.9 only considers the case of linear constraints, but the constraints in this problem are actually linear.)

4.2 In this exercise we consider the left- and right-hand sides of (4.22) for the case where the feasible set of the primal problem is $\mathbb{S} = \{x \in \mathbb{R}^n | g(x) = \mathbf{0}\}$. That is, we only have equality constraints and we can neglect the dual variables μ corresponding to the inequality constraints.

(i) Consider the primal problem $\min_{x \in \mathbb{R}^2}\{f(x)|g(x) = 0\}$ where the functions $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ are defined as:

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

Evaluate the left- and right-hand sides of (4.22) for this f and g. That is, evaluate $\min_{x \in \mathbb{R}^2}\{f(x)|g(x) = 0\}$ and $\sup_{\lambda \in \mathbb{E}} \mathcal{D}(\lambda)$. Be careful that you actually find an infimum of the inner problem. Is there a duality gap?

(ii) Repeat the previous part but re-define f to be:

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

(iii) Repeat the previous part but re-define f to be:

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

4.3 Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by:

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

(i) Calculate ∇f.
(ii) Calculate $\nabla^2 f$.
(iii) Show that f is convex.
(iv) Show that no x exists satisfying $\nabla f(x) = 0$.
(v) Show that there is no minimizer of $\min_{x \in \mathbb{R}} f(x)$.

4.4 In this exercise we use MATLAB to minimize two functions.

(i) Use the MATLAB function `fminunc` to minimize $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $\forall x \in \mathbb{R}^2, f(x) = (x_1 - 1)^2 + (x_2 - 3)^2 - 1.8(x_1 - 1)(x_2 - 3)$.
You should write a MATLAB M-file to evaluate both f and ∇f. Specify that you are supplying the gradient ∇f by setting the `GradObj` option to `on`. Set the `LargeScale` option to `off`. Use initial guess $x^{(0)} = \begin{bmatrix} 3 \\ -5 \end{bmatrix}$. Report the number of iterations required.

(ii) Repeat the first part, but minimize the function $f : \mathbb{R}^4 \to \mathbb{R}$ defined by:

$$\begin{aligned}\forall x \in \mathbb{R}^4, f(x) = {} & (x_1 - 1)^2 + 2(x_2 - 3)^2 + 2(x_3 - 1)^2 + (x_4 - 3)^2 \\ & - 1.8(x_1 - 1)(x_2 - 3) - 1.8(x_2 - 3)(x_3 - 1) - 1.8(x_3 - 1)(x_4 - 3),\end{aligned}$$

using initial guess $x^{(0)} = \begin{bmatrix} 3 \\ -5 \\ 3 \\ -5 \end{bmatrix}$. Report the number of iterations required.

4.5 Consider the problem $\min_{x \in \mathbb{R}^2} \{ f(x) | Ax = b \}$ where $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by:

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

with:

$$Q = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, c = \begin{bmatrix} 4 \\ 3 \end{bmatrix},$$

and the coefficient matrix and right-hand side of the constraints is specified by:

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

(i) Solve the problem by solving the first-order necessary conditions.

(ii) Use the MATLAB function `quadprog` to solve the problem. Use initial guess $x^{(0)} = \begin{bmatrix} 3 \\ -5 \end{bmatrix}$.

4.6 Consider Problem (4.5) from Section 4.1.8:

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

where $f : \mathbb{R}^2 \to \mathbb{R}$, $A \in \mathbb{R}^{1\times 2}$, and $b \in \mathbb{R}^1$ were defined by:

$$\begin{aligned} \forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\ A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\ b &= \begin{bmatrix} 0 \end{bmatrix}. \end{aligned}$$

Suppose that the equality constraints changed from $Ax = b$ to $Ax = b - \gamma$.

(i) Calculate the sensitivity of the minimum to γ, evaluated at $\gamma = [0]$.

(ii) Solve the changed problem explicitly for $\gamma = [0.1]$ and compare to the estimate provided by the sensitivity analysis.

(iii) Repeat the previous part for $\gamma = [1]$.

4.7 Consider Problem (4.10):

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

where

$$\begin{aligned}
\forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\
A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\
b &= \begin{bmatrix} 0 \end{bmatrix}, \\
C &= \begin{bmatrix} 0 & -1 \end{bmatrix}, \\
d &= \begin{bmatrix} -3 \end{bmatrix}.
\end{aligned}$$

(i) Use the MATLAB function `quadprog` to find the minimizer and minimum of the problem. Use as initial guess $x_1^{(0)} = 5, x_2^{(0)} = 5$.
(ii) Form the dual of the problem.
(iii) Use the MATLAB function `quadprog` to find the maximum of the dual problem. Use as initial guess $\mu^{(0)} = [0.25], \lambda^{(0)} = [0]$.

4.8 Consider Problem (4.10), which has objective $f : \mathbb{R}^2 \to \mathbb{R}$ and equality constraint $Ax = b$ defined by:

$$\begin{aligned} \forall x \in \mathbb{R}^2, f(x) &= (x_1 - 1)^2 + (x_2 - 3)^2, \\ A &= \begin{bmatrix} 1 & -1 \end{bmatrix}, \\ b &= \begin{bmatrix} 0 \end{bmatrix}. \end{aligned}$$

However, suppose that the inequality constraint was changed to $Cx \leq d - \eta$, with $C \in \mathbb{R}^{1\times 2}$ and $d \in \mathbb{R}^1$ defined by:

$$\begin{aligned} C &= \begin{bmatrix} 0 & -1 \end{bmatrix}, \\ d &= \begin{bmatrix} -3 \end{bmatrix}. \end{aligned}$$

Let $\eta = [0.1]$.

(i) Use Theorem 4.14 to estimate the change in the minimum due to the change in the inequality constraint.

(ii) Solve the change-case problem explicitly and compare the result to that obtained by sensitivity analysis.